

THIS WEEK



EDITORIALS

PUBLIC HEALTH A questionable 'epidemic' of chronic diseases **p.250**

WORLD VIEW Advising on public safety without ending up in court **p.251**

TAKING ITS TIME The 47-hour day of the blind cavefish **p.253**

A paler shade of green

The Obama administration should reject the false dichotomy between environmental protection and the economy.

Even at the best of times, building political support for action on thorny environmental issues is difficult in the United States. Recent events serve as a stark reminder of what happens when times are hard. Faced with the alarming possibility of a double-dip recession and an energized opposition that has demonized environmental regulations of any kind, President Barack Obama is picking his battles carefully and seeking to carve out a middle ground on what will surely be the fundamental issue of the 2012 presidential election: the economy. Sadly, if understandably, the environment has been placed on the back burner.

An 8 September speech by Obama calling for legislation to reinvigorate a stalled economic recovery marked a striking shift from the economic stimulus bill of 2009, a time of easy rhetoric about 'green jobs'. This time around, global warming didn't even get a mention — and neither did clean energy, once hailed as the basis for sustainable growth and global economic competitiveness. The lamentable truth is that in the world of US politics, environmental protection is still debated as if it were an optional and expensive accessory to modern living. In the process, science is set aside.

The latest setback came on 2 September, when Obama ordered Environmental Protection Agency (EPA) administrator Lisa Jackson to stand down on tightened standards for ozone pollution. Jackson had been working to plug a hole left by the administration of former President George W. Bush, which in 2008 set weaker standards than recommended by the EPA's science advisers. Obama elected to opt out of the current rule and promised to push forward with another ozone review that is due out in 2013.

The idea that tightening ozone standards would damage the economy is questionable at best. Numerous studies have shown that pollution control tends to pay for itself by reducing public-health bills; moreover, money spent on reducing emissions does not disappear into a vacuum: pollution control is a business, too. In backing off from the tighter regulations, Obama was looking to disarm his political opponents more than anything else, but in doing so he lent false legitimacy to the misguided debate that pits the economy against public health and environmental protection.

The Obama administration has also been cautious in extending its regulatory powers to the overwhelming environmental issue of our time, climate change. Yet regulation based on existing laws is the only remaining tool for addressing the issue now that greenhouse-gas legislation is off the table. A proposed pipeline from the Canadian tar sands in Alberta to US refineries along the Gulf of Mexico has become the latest flashpoint for disputes over climate policy. Activists say that the pipe will accelerate the extraction of oil from the tar sands, and hundreds of people — including NASA climate scientist James Hansen — were arrested at protests staged at the White House in late August.

Final word on the Keystone XL pipeline rests with the Department of State, which issued an environmental-impact statement on the

project last month and is widely expected to approve the project this autumn. In fact, the pipeline protests say more about the sorry state of the environmental agenda than anything else. It is true that greenhouse-gas emissions from oil extracted from the sands are 15–20% higher than those from average crude oil if assessed on a life-cycle

"Money spent on reducing emissions does not disappear into a vacuum."

basis, but industry officials are correct in pointing out that this is on a par with other dirty oils produced in the United States and elsewhere using steam injection. And halting this pipeline is unlikely to halt development of the tar sands or other dirty sources of energy. What is missing, now as ever, is a

policy to address the larger climate threat.

Science-based regulation may yet have a chance. The EPA will soon propose regulations that would reduce emissions of mercury and other pollutants from power plants and other industrial sources. In his address to Congress, Obama insisted that the economic crisis should not be used as an excuse to wipe out basic economic, health and environmental protections. "We shouldn't be in a race to the bottom," Obama said, "where we try to offer the cheapest labour and the worst pollution standards." Obama and his administration still have the opportunity to live up to those words. ■

Patent medicine

A simplification of the US patent system is good news for inventors, but could have gone further.

The passage of a patent-reform bill by the US Senate on 8 September was a rare win for President Barack Obama, who on the same day gave a high-profile speech on job creation and argued that patent reform was part of the solution. The America Invents Act, as it is called, is also good news for researchers and their institutions.

The link to jobs is speculative, but the bill is likely to simplify life for inventors. Most significantly, it moves the United States to a first-to-file system, in which patents are granted to those who get their applications to the patent office first. That should eliminate the lengthy administrative procedures that are often required to determine who has the true priority on inventions under the current first-to-invent system. Any scientist who has ever been caught up in a patent wrangle — such as the competition between Bell Laboratories and IBM for the US patent on the high-temperature superconductor yttrium barium copper oxide, which famously took 13 years to be settled in favour of Bell Labs — will see the advantages of that. ►

► The bill is also expected to reduce costly patent litigation by ensuring, through a review procedure after the patent is granted, that all patents describe working inventions. ‘Patent trolls’ — individuals and companies who attempt to make money by filing broad patents, then suing those who use similar technologies — will not welcome the move, but legitimate inventors will. A third important change is a presumption that the US Patent and Trademark Office will, at Congress’s discretion, be able to keep the filing fees that it raises each year, rather than see them diverted to other parts of the government. That will leave the office with extra resources to clear its backlog of patent applications.

An earlier version of the bill would have prevented Congress from diverting the funds at will, but that provision was watered down by politicians keen to retain congressional control of the budget. The bill also misses an opportunity to loosen constraints placed on research and medicine by gene patents. Researchers or companies who independently develop diagnostic tests based on genes that have already been patented risk being sued for patent infringement. In July, a New York appeals court underscored that risk when it upheld the rights of Myriad Genetics, a genetic-testing company in Salt Lake City, Utah, to enforce its patents on genes implicated in breast and ovarian cancer.

As early as 2006, a National Academy of Sciences panel recommended that Congress consider an exception to the enforceability of patents on genes used for diagnostic tests, to allow independent confirmation of the results. And the Secretary’s Advisory Committee on Genetics,

“Gene patents stifle research and restrict patients’ access to second opinions.”

Health, and Society — a panel convened by the Department of Health and Human Services — found in 2010 that gene patents were stifling research and restricting patients’ access to second opinions. The committee strongly recommended exemptions for anyone conducting independent tests or basic research.

An amendment to the America Invents Act could have implemented such exemptions, but now the bill merely calls for yet another study of the issue.

Still, the bill’s passage with bipartisan support is a precious exception to the polarization that has characterized US political debates as campaigns for the 2012 presidential election get under way. Given that attempts to update the US patent system have failed repeatedly in recent years, researchers should be happy to see reform implemented at last. ■

Disease priorities

Non-communicable diseases are on the rise. Emerging nations need to take them seriously.

Although much of the world’s population remains poor, most people die from diseases once associated with wealth: cancer, heart disease, diabetes, to name a few. Next week in New York, a high-level summit run by the United Nations will put the threat of such non-communicable diseases (NCDs) firmly on the international agenda. The summit is likely to yield few surprises. A draft of the political declaration, agreed on last week, is short on specific proposals and postpones debate on controversial measures such as mandatory salt reductions in foods. But by recognizing the threat of NCDs and pledging action, it marks a victory for public-health experts and non-governmental organizations, which have long argued that in a world of emerging economies and successful campaigns against infectious diseases, it is time to tackle what many call an ‘epidemic’ of NCDs in poorer nations.

The sheer number of deaths from NCDs — 36 million in 2008, or 63% of all deaths worldwide — certainly suggests that these diseases should share the global health agenda with communicable diseases, such as malaria, AIDS and tuberculosis. In fact, infectious diseases have lit the way forward for NCDs by showing how to put diseases of the poor on the political agenda and vastly increase support for their control. But they have their limits as a metaphor for the challenge of NCDs. In key respects, both the problem and potential solutions are very different.

There’s no question that these diseases are a big problem in poorer nations, but how big is far from clear (see page 260). The World Health Organization and disease groups have a tendency to emphasize headline-grabbing figures. But the number of deaths from NCDs does not tell the full story. More to the point is the age at which a disease strikes and, therefore, the years of life that it steals. On that score, infectious diseases remain a much bigger burden, at least in poorer countries; HIV/AIDS continues to wreak devastation in sub-Saharan Africa.

Talk of an epidemic of NCDs also omits the fact that in poorer countries, such diseases are driven more by demographic changes than by behavioural factors such as obesity and smoking. In many poor and middle-income countries, the age structure of the population is changing, as birth rates fall and the large number of people

born in past decades enter middle age. As a result, more people are likely to develop an NCD, which mostly affect older people.

How these trends will play out is unclear. Projections of NCD mortality are too often accepted without question, even though they are based on rudimentary models that rely on patchy data from poorer countries, combined with historical trends on the incidence of ailments in wealthier countries and simple parameters such as expected GDP growth. Projections of NCD mortality in poorer countries should be treated with healthy caution. The summit’s call for a way to improve monitoring and data collection on NCDs is therefore welcome and long overdue.

Some of the countries where NCDs are now surging have an opportunity to control them much more quickly than wealthier countries have in the past. In sharp contrast with infectious diseases, which were often neglected, NCDs have been the prime focus of almost all biomedical research and drug development in rich countries. An array of drugs and technologies already exists, as do decades of best practice. Emerging economies could leapfrog richer countries by tapping into these advances.

Poorer countries could also vastly expand access to the many cheap drugs already available there for NCD control, such as statins and aspirin for heart disease, and by implementing well-understood and effective prevention measures to reduce risk factors, such as banning smoking in the workplaces, bars and restaurants. The international community can help by designing public-health programmes and ensuring a supply of inexpensive drugs.

In another respect, NCDs are a much less tractable problem than infectious diseases, and are less well suited to international intervention. When donors commit funds for a vaccination campaign, for example, they can confidently predict the number of lives their investment will save. Efforts to control infectious diseases tend to be faster and simpler than the more drawn-out and complex task of treating chronic diseases, and of addressing risk factors such as obesity and smoking.

Ultimately, a sustained assault on NCDs will require strong national health systems. There the international community can at best catalyse and help to shape efforts. But only national governments can sustainably fund the bulk of health infrastructure and staff. It can only be hoped that the new international focus will spur emerging economies and poorer nations to give NCDs more attention in their funding prioritizations. International summits can help, but it will be down to national governments to really make a difference. ■

► **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xbhunq

N. TUCKER



Check your legal position before advising others

Next week's trial of seismologists in Italy highlights the risks to scientists who offer public advice. **Willy Aspinall** considers what can be done.

The world is litigious and scientists are not immune. Next week, six scientists and an official are scheduled to go on trial in Italy charged with multiple manslaughter. Their alleged crime? That they were negligent in giving advice on the risk to public safety during the seismic unrest that culminated in a magnitude-6.3 earthquake near L'Aquila in central Italy on 6 April 2009, which killed more than 300 people. Prosecutors in Italy say that residents were misinformed by the group's advice, and that this contributed to some people choosing not to leave their homes, with fatal consequences (see page 264).

This is not a trial of earthquake-prediction science, as some seismologists seem to think. Rather, it is about possible negligence in the provision of hazard-assessment advice, for which there is little or no case law or precedent, unlike, say, professional liability in civil engineering or medicine.

Even before the trial begins, consequences of such legal action are clear: knowledgeable scientists may distance themselves, leaving those who are largely naive, dogmatic or blasé about legal risks to offer opinions.

I have personal experience of these issues. In 1997, I was chief scientist at the Montserrat Volcano Observatory in the Caribbean when an eruption of the Soufrière Hills volcano killed 19 people. After the eruption, a scientific advisory panel was set up, of which I am a member, to issue outlooks every six months or so. Together with day-to-day advice from the observatory, these alerts underpin policy decisions, such as where entry is allowed and where is off-limits.

We saw our work as a civic responsibility, but our legal position was uncertain, and eventually it became clear that we might be vulnerable to claims — civil or criminal, genuine or vexatious. One Montserrat resident who was injured in the eruption filed a lawsuit against government officials for failing to enforce protective measures; others sued on the basis that those same measures had infringed their rights of access to their homes. The government countered that it had acted lawfully and, pertinent for us, on scientific advice. However, the nature and legal standing of our advice were never tested in court.

We felt our position was clarified when the UK government, responsible for security and public safety on Montserrat as an overseas territory, issued guidelines on scientific advice in 2001 (updated in 2007). These included a clause that seems to indemnify committee members, provided that they have “acted honestly, reasonably, in good faith and without negligence”. However, under UK law, negligence can be decided only in court, so this cannot deflect action under all circumstances.

It is worth mentioning that volcanologists are more familiar with short-term scenario forecasting than seismologists, who tend to concentrate

on advancing theoretical understanding. Short-term ‘operational earthquake forecasting’ — using activity traits to infer increases in earthquake hazard level — has been initiated in California to alert people after a big quake to the immediacy and size of aftershocks. But strong resistance to the concept remains. Notorious failed earthquake predictions from the 1970s have left many seismologists hesitant about the notion, concerned that it is prediction in another guise. This scientific caution can, arguably, make us unreceptive to hints of an impending threat, as, for example, with the unusual sequence of quakes that occurred off the coast of Japan on 9 March 2011, two days before the disastrous quake and tsunami.

What is to be done? It is always difficult to convey scientific uncertainty without giving the impression that nothing useful is known, but overstating scientific certainties can be more dangerous. Volcanologists

have adopted a protocol on professional conduct in crises (see go.nature.com/wjueqm), and some of the principles could be helpful to seismologists for situations such as L'Aquila.

Certainly, scientists who provide assessments and forecasts must be aware of legal implications. Ideally, they should provide advice in writing, staying within their domain of expertise and citing evidence that is robust under peer review and defensible in law. Sloppy arguments and casual errors — even in reports or papers elsewhere — risk exposure if a related issue crops up in subsequent legal proceedings.

If verbal advice must be given, scientists should make a record of it — public officials on the receiving end are certain to keep notes. From

experience, critical phone calls during a crisis should be recorded: even the precise timing of a call could be material in retrospect. Off-the-cuff comments are easily misconstrued, sometimes wilfully, so scientists in sensitive situations should think carefully about their use of social media. Electronic messaging can propagate alerts — and rumours — instantly and widely, but the legal status of their content remains unclear.

One change could be that the same level of legal liability protection granted to colleagues such as weather forecasters in federal or national agencies is afforded to scientists in official advisory roles. When the lives of thousands of people are at risk in a crisis, university and independent specialists often work pro bono. It is more than poignant that resources for providing scientific advice before a disaster are invariably dwarfed by those devoted to scrutinizing that advice in a legal post-mortem. And it is salutary that scientists who have shouldered professional obligations voluntarily can find themselves legally exposed. ■

Willy Aspinall is Cabot professor in natural hazards and risk science at Bristol University, UK.
e-mail: willy@aspinall.demon.co.uk

**SCIENTISTS IN
SENSITIVE
SITUATIONS SHOULD
THINK CAREFULLY
ABOUT THEIR USE OF
SOCIAL MEDIA.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/ld8icr

SEVEN DAYS

The news in brief

POLICY

GM honey concern

The European Court of Justice ruled on 6 September that honey containing trace amounts of pollen from genetically modified (GM) plants can no longer be sold in the European Union without a safety review. The decision could affect imports of honey from countries such as Argentina, which are big producers of GM crops, and increase the already powerful resistance to the cultivation of GM crops in Europe.

Blood donations

Like Australia, Sweden and Japan, Britain will allow gay men to give blood, as long as their last sexual contact with another man was more than 12 months ago. The Advisory Committee on the Safety of Blood, Tissues and Organs said on 8 September that a lifetime ban on blood donations — introduced in the 1980s — could be lifted (from 7 November) without increasing the risk of infections such as HIV being transmitted. The United States and Canada still prohibit gay men from giving blood.

Ozone network cut

As part of cuts due to budget constraints, Canada's environment agency has decided to shut down a network of stations that monitor ozone levels in the Arctic. Environment Canada's ozone and radiation research group will also be substantially reduced, and the agency has said that it will no longer host a long-running archive of ozone data. See page 257 for more.

US patent reform

US inventors should find it easier to get and defend patents following the long-awaited passing of reforms

to the patent system. The America Invents Act — passed by the Senate on 8 September, after clearing the House of Representatives in June (see *Nature* **472**, 149; 2011) — will switch the US system from one where considerable resources are spent establishing priority for inventions, to a system in which the patent is granted to whoever files an application first. See page 249 for more.

Illegal fishing

The United States and Europe have pledged greater cooperation in fighting illegal fishing, which has pushed many species into decline. Jane Lubchenco, head of the US National Oceanic and Atmospheric Administration, and Maria Damanaki, the European Union fisheries commissioner, made the

announcement last week, but gave few details beyond references to improving monitoring and enforcement.

RESEARCH

Exoplanet trove

European astronomers announced on 12 September the discovery of more than 50 exoplanets, including one that sits in the 'habitable zone': the distance at which it orbits its star means that the planet could harbour liquid water. The discoveries were made with the High Accuracy Radial Velocity Planet Searcher (HARPS) on a telescope at La Silla Observatory in Chile, run by the European Southern Observatory. They are the latest salvo in the competition between ground-based telescope teams (such as

HARPS) and the NASA space telescope, Kepler. See go.nature.com/l9cptom for more.

Cancer lawsuits

Eight patients with cancer who were enrolled in clinical trials based on faulty research are suing the scientists involved and their institution, Duke University in Durham, North Carolina. The lawsuit, filed on 7 September, relates to trials based on the work of cancer geneticist Anil Potti, who claimed to have found links between patients' gene-expression profiles and their response to chemotherapy drugs. After other scientists raised questions about the analysis, Potti resigned and five papers were retracted; the trials are now suspended (see *Nature* **469**, 139–140; 2011). But the plaintiffs allege that researchers



K. HANSEN/NASA

Arctic sea ice drops to record low

The extent of sea ice in the Arctic dropped to a new record minimum last week, according to researchers at the University of Bremen in Germany, who used high-resolution microwave data from a sensor on board NASA's Aqua satellite. At 4.24 million square kilometres, sea-ice cover on 8 September was 27,000 square kilometres smaller than

the previous record low, observed in 2007. Melting on the surface of the ice (pictured, during a NASA ICESCAPE research mission in July) has already ended, but late-season ice loss may continue, say scientists with the US National Snow and Ice Data Center, in Boulder, Colorado, as warm water continues to melt the ice from below.

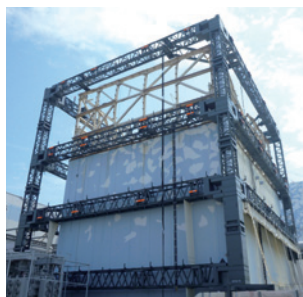
TECO and officials at Duke University pushed forward with the trials, despite knowing that the research was flawed.

Mapping the Moon

Two spacecraft that will fly in tandem around the Moon to precisely map its gravitational field, and therefore the composition of its interior, were launched from Cape Canaveral, Florida, on 10 September. NASA's Gravity Recovery and Interior Laboratory (GRAIL) mission will take a circuitous 3.5-month journey to lunar orbit and will not start to collect data until early March 2012. For more on the mission, see *Nature* 477, 16–17 (2011).

Cancer centre cut

As part of a major shake-up, the Ludwig Institute for Cancer Research (LICR) is closing its colorectal-cancer centre in Parkville, Australia. Scientists there say that they are disappointed and bewildered by the decision. The LICR, a non-profit organization, spends more than US\$100 million each year at ten centres around the world and employs more than 700 staff, focusing on translating basic research into novel cancer therapies. But it wants to concentrate research at two or three large hubs and close down small branches. See go.nature.com/fy8xky for more.



EVENTS

Fukushima cover

Six months after the meltdowns at the Fukushima Daiichi nuclear power plant in Japan, workers have completed a metal frame over the damaged unit 1 reactor (pictured), ready to hold a cover to shield it from the wind and rain and lessen the chance of radioactivity spreading. Workers continue to pour water into the reactor core, which is currently at 85–90 °C. Meanwhile, Japan's trade and industry minister, Yoshio Hachiro, resigned just 9 days into his post after local media reported that he had referred to an exclusion zone near the plant as a “town of death”, and joked about radiation at a press conference.

Nuclear explosion

One worker was killed and four injured by a furnace explosion at a facility for incinerating low-level nuclear waste near Codolet, southern

France, on 12 September. The plant, known as CENTRACO, is administered by a subsidiary of the French energy company EDF. A spokeswoman said that no radioactivity had been released beyond the site. See go.nature.com/ixhoef for more.

PEOPLE

NASA spy jailed

A former NASA scientist has been sentenced to 13 years in prison after admitting that he tried to sell classified information to Israel. Stewart Nozette was a distinguished expert on defence and space technology, and principal investigator for the radar instrument on NASA's Lunar Reconnaissance Orbiter. He worked for a series of US government agencies in the 1980s and 1990s. He was arrested in October 2009 after an FBI sting operation, and made a plea deal with prosecutors on 7 September. See go.nature.com/dra37v for more.

Suspected fraud

Tilburg University in the Netherlands announced on 7 September that it had suspended Diederik Stapel, a prominent Dutch psychology professor, because he had used ‘fictitious’ data in his research. Stapel, 45, was director of the Tilburg Institute for

COMING UP

19–20 SEPTEMBER

In New York, the United Nations holds a major summit on tackling non-communicable diseases such as heart attack and cancer. See also page 260. go.nature.com/oc9r7t

19–21 SEPTEMBER

At Google's headquarters in Mountain View, California, the US National Academy of Sciences hosts a meeting on the ‘frontiers of engineering’, including sessions on neuroprosthetics and sustainable buildings. go.nature.com/qfpz1h

Behavioral Economics Research, and his work explored power, stereotyping and other social behaviours. A committee chaired by a former president of the Royal Netherlands Academy of Arts and Sciences will scrutinize Stapel's work at Tilburg, where he joined the faculty in 2006, and at his previous institution the University of Groningen, and publish its findings by the end of October.

Lasker award

The US\$250,000 Basic Medical Research Award — whose winners often go on to receive a Nobel prize — was this year awarded to two protein biochemists: Franz-Ulrich Hartl of the Max Planck Institute for Biochemistry in Munich, Germany, and Arthur Horwich of Yale University School of Medicine in New Haven, Connecticut. The duo helped to establish how proteins called chaperonins assist other proteins in folding into complicated three-dimensional shapes.

➔ NATURE.COM

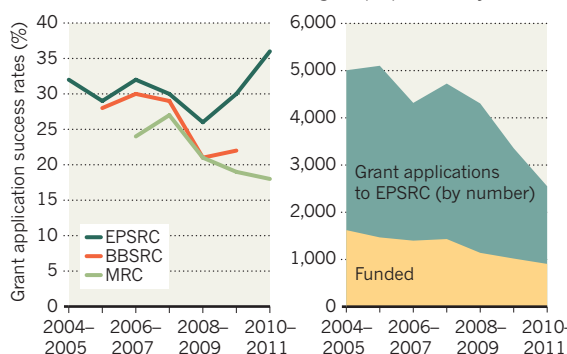
For daily news updates see: www.nature.com/news

TREND WATCH

Britain's Engineering and Physical Sciences Research Council seems to be reaping the benefits of its sometimes controversial efforts to reduce the number of grant applications it receives (see *Nature* 464, 474–475; 2010). In 2010–11, the agency funded 912 of 2,568 applications: a smaller number than in previous years, but at 36% a much greater success rate (see charts). By contrast, success rates at other UK research funding agencies are still falling. See go.nature.com/tlmcij for more.

SUCCEEDING IN BRITAIN

By cutting applications, the Engineering and Physical Sciences Research Council funded more than one in three grant proposals last year.



EPSRC: Engineering and Physical Sciences Research Council; BBSRC: Biotechnology and Biological Sciences Research Council; MRC: Medical Research Council.

NEWS IN FOCUS

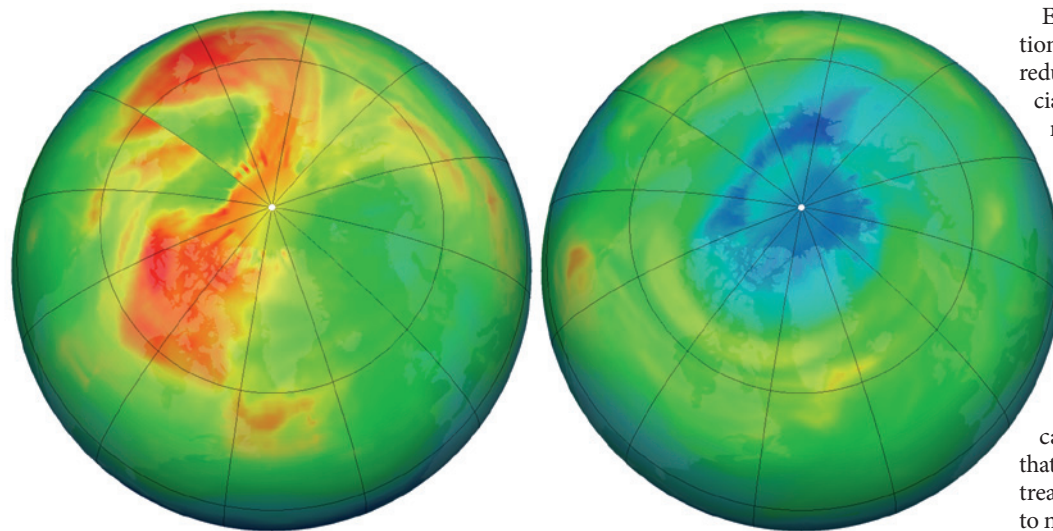
COMMUNITY Paul Nurse reveals plans to shake up Britain's Royal Society **p.258**

PLANT SCIENCE Weed researchers, meet crop breeders **p.259**

ENVIRONMENT United Nations seeks a definition of 'desertification' **p.262**



SEISMOLOGY How six scientists came to be tried for manslaughter **p.264**



Arctic ozone levels hit a record low this year (blue area, right), compared with a relative high (red) in 2010.

ENVIRONMENT

Canadian ozone network faces axe

Arctic monitoring stations hit by budget constraints.

BY QUIRIN SCHIERMEIER

A key source of information about the health of the ozone layer above the Arctic looks set to be choked off.

In a year that saw the first genuine 'ozone hole' appear in the Northern Hemisphere, atmospheric scientists say they are shocked to learn that Environment Canada, the country's environment agency, has decided to drastically reduce its ozone science and monitoring programme.

Its network of monitoring stations provides about one-third of the Arctic's ozone measurements and this year contributed key data showing unprecedented depletion of stratospheric ozone over the Arctic. With regular *in situ* measurements going back to 1966, Canada also holds the longest-running record of atmospheric ozone levels in the

world — an archive that is also threatened.

The Canadian observation network comprises 17 stations — from London, Ontario, in the south to Alert in the high Arctic — which use several techniques to measure ozone (see 'The ozone network'). But atmospheric scientists and research institutes around the world, including Canada, Britain, Switzerland and Germany, have been told, informally, that the network will be shut down as early as this coming winter. This will be the end of *in situ* ozone measurements, including those made by balloons launched at least once a week from 11 of the stations. "This is devastating for the whole field," says Tom Duck, who conducts atmospheric research at Dalhousie University in Halifax, Nova Scotia.

➔ **NATURE.COM**
For more on the science of the ozone layer see:
go.nature.com/2xjzjcc

Environment Canada's ozone and radiation research group will also be substantially reduced as a result of staff cuts driven by financial constraints. A spokesman for the agency refused to confirm the cuts, saying merely that all government-funded programmes are currently being reviewed.

"This is a sad and abrupt end of many years of very successful collaboration," says Markus Rex, an ozone researcher at the Alfred Wegener Institute for Polar and Marine Research in Potsdam, Germany, who has previously worked with the Canadian network in international, Arctic-wide ozone measurement campaigns, dubbed Match. "I'm worried that the cuts will lead to an erosion of the Montreal Protocol, which obliges Arctic countries to monitor the ozone layer and maintain scientific ozone research."

The blow comes at a crucial time for monitoring efforts. In March, scientists reported that 40% of stratospheric ozone over the Arctic had been destroyed — the highest ozone loss previously measured was 30% in 2005. The record loss was due mainly to exceptionally low temperatures last winter in the Arctic stratosphere, which help to form ice particles at an altitude of around 18–25 kilometres. These particles host the chemical reactions by which long-lived chlorofluorocarbons catalyse the breakdown of ozone.

Although ozone loss is usually more pronounced above cooler Antarctica, ozone-depleted Arctic air can reach populated areas more easily. Because ozone helps to block potentially harmful ultraviolet radiation from the Sun, its depletion could raise the risk of health problems such as skin cancer in northern populations.

"Canada has been a linchpin of Arctic ozone observation," says Neil Harris, an atmospheric chemist at the University of Cambridge, UK, who heads the European Ozone Research Coordinating Unit. "It has contributed very substantial data to research that allows us to be diagnostic about what's happening in the Arctic stratosphere. If we were to lose one-third of our monitoring capability in the Arctic the overall loss in scientific value will be much greater."

Satellite observations are no substitute for *in situ* ozone measurements, adds Johannes Staehelin, an atmosphere researcher at the Swiss Federal Institute of Technology in ▶

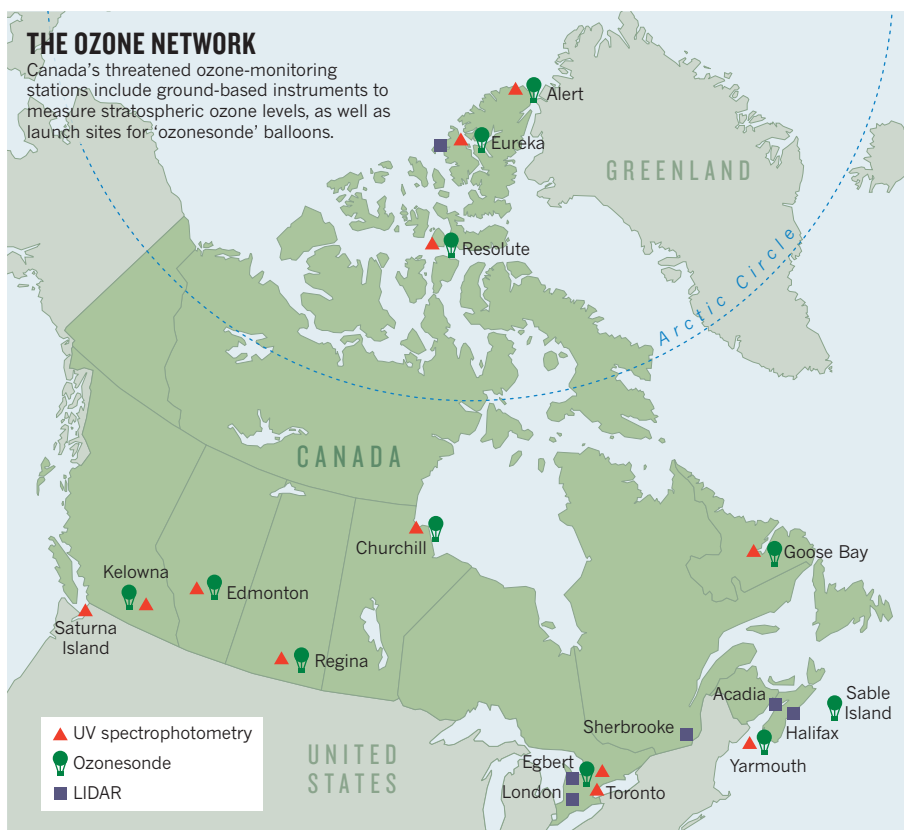
R. SIMMON/NASA

► Zurich who chairs the World Meteorological Organization's ozone science advisory group. Indeed, *in situ* data are essential for calibrating and validating measurements by satellites such as NASA's Aura and the European Space Agency's Envisat.

Staehelin adds that the Canadian agency has said it will no longer host the Toronto-based World Ozone and Ultraviolet Radiation Data Centre, an archive of data collected over several decades and used intensively by atmospheric scientists around the world. "It appears that the management at Environment Canada was not fully aware of the consequences of its decision," says Staehelin. Last month, the agency notified its staff that a total of about 300 jobs will be cut.

Canadian environmental research has already been hit hard by the looming closure of the Canadian Foundation for Climate and Atmospheric Sciences, which provides the majority of funds for climate and atmospheric science in the country. The charitable foundation has received no federal funding since 2003, and is expected to close next year.

"The funding crisis in this country is really hammering our ability to observe and protect the environment of Canada," says Duck. "I have already lost most of my group because I just can't pay them any more. If help doesn't come soon, many others will shut up shop." ■



SCIENTIFIC SOCIETIES

Nurse takes Royal Society's pulse

President plans wider role for Britain's national academy.

BY GEOFF BRUMFIEL

For Paul Nurse, few things are off-limits. Since taking over as president of Britain's Royal Society in December last year, he has been overseeing a strategic review that is likely to lead to the first change to the society's charter since it was signed by King Charles II in 1662.

The change is relatively minor (it extends the terms of office for the society's council members), but it gives a good indication of how he is likely to approach his five-year tenure. "I felt we should look at everything we do, root and branch," he says over morning tea in the society's august central-London headquarters.

Nurse wants the society to have a stronger voice on the big policy questions of the day. "The Royal Society has a responsibility to provide advice on difficult issues, even if they are contentious," he says.

He hopes to boost the society's role in government decision-making by fostering greater involvement of its roughly 1,500 fellows and foreign members in preparing reports,

potentially with the help of more policy staff. Nurse also wants to expand the number of authoritative and influential reports on key issues, such as nuclear power, climate change and the definition of life. The society has long produced such reports, most recently on the global scientific enterprise and on the potential threats and opportunities offered by geoengineering to mitigate climate change. But Nurse sees an opportunity to do more on a broader range of topics, with an eye for increasing the society's global reach. "I think the world would listen to us," he says.

Not everyone is convinced. "The first thing it should do is get a big bookshelf and put it in the basement to store the reports," says Daniel Greenberg, a journalist based in Washington DC who has devoted his career to studying the intersection of science and policy. The US National Academy of Sciences, which produces many more reports than its British counterpart, has relatively little influence

over the political process, he says. "Nobody in politics reads an academic report, slaps the side of their head and says 'Wow!'," he says.

Others suggest that the society could gain more influence by choosing its topics carefully. National Academy reports can sometimes shift the tone of a debate, says David Goldston, who was chief of staff on the US House Committee on Science from 2001 to 2006. And Robert May, a zoologist at the University of Oxford, UK, and former president of the society, points out that a 2002 report on foot-and-mouth disease helped to set up national vaccination strategies to prevent widespread cattle culling during future outbreaks. That report was successful in part because the society consulted closely with politicians and bureaucrats throughout, says May.

Nurse insists that the society will focus on "big areas that are important to our society" — not just those immediately relevant to policy-makers. Through conferences and studies, the society should also draw attention to the biggest mysteries in science, he says. "What is life? What is the beginning of the Universe? You know, that type of question." ■

➔ **NATURE.COM**
To read more about the history of the Royal Society, see: go.nature.com/ythctz

CORRECTION

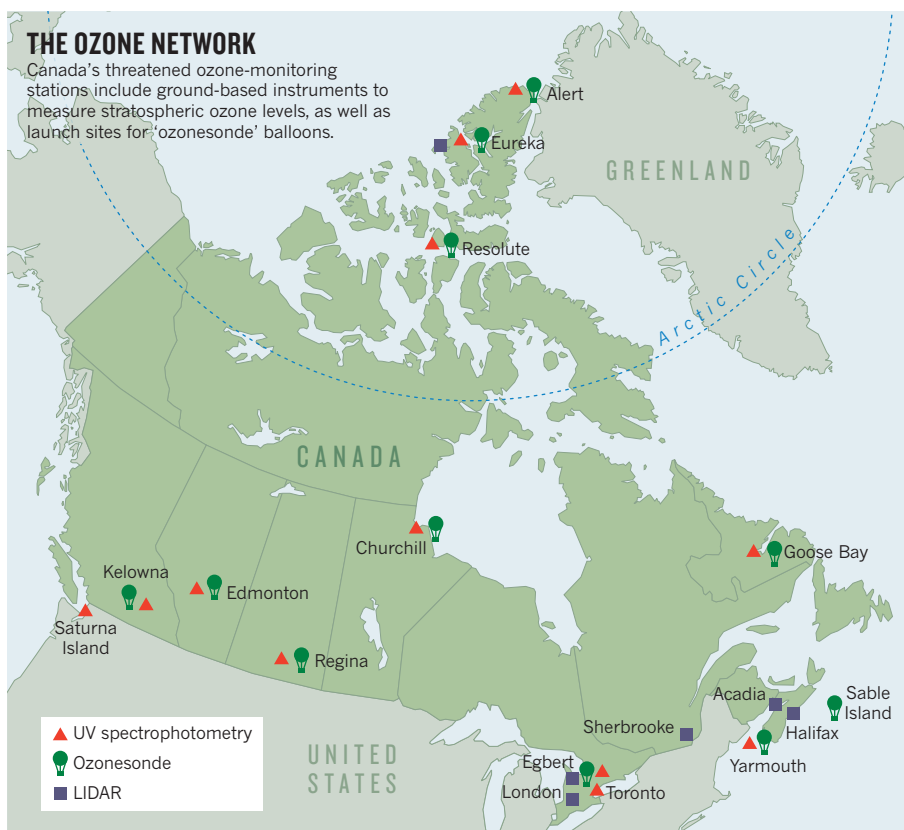
The news story 'Canadian ozone network faces axe' (*Nature* **477**, 257–258; 2011) originally stated that Environment Canada planned to cut 776 jobs. Although 776 employees will be affected by workforce changes, only about 300 posts are being eliminated.

► Zurich who chairs the World Meteorological Organization's ozone science advisory group. Indeed, *in situ* data are essential for calibrating and validating measurements by satellites such as NASA's Aura and the European Space Agency's Envisat.

Staehelin adds that the Canadian agency has said it will no longer host the Toronto-based World Ozone and Ultraviolet Radiation Data Centre, an archive of data collected over several decades and used intensively by atmospheric scientists around the world. "It appears that the management at Environment Canada was not fully aware of the consequences of its decision," says Staehelin. Last month, the agency notified its staff that a total of 776 jobs will be cut.

Canadian environmental research has already been hit hard by the looming closure of the Canadian Foundation for Climate and Atmospheric Sciences, which provides the majority of funds for climate and atmospheric science in the country. The charitable foundation has received no federal funding since 2003, and is expected to close next year.

"The funding crisis in this country is really hammering our ability to observe and protect the environment of Canada," says Duck. "I have already lost most of my group because I just can't pay them any more. If help doesn't come soon, many others will shut up shop." ■



SCIENTIFIC SOCIETIES

Nurse takes Royal Society's pulse

President plans wider role for Britain's national academy.

BY GEOFF BRUMFIEL

For Paul Nurse, few things are off-limits. Since taking over as president of Britain's Royal Society in December last year, he has been overseeing a strategic review that is likely to lead to the first change to the society's charter since it was signed by King Charles II in 1662.

The change is relatively minor (it extends the terms of office for the society's council members), but it gives a good indication of how he is likely to approach his five-year tenure. "I felt we should look at everything we do, root and branch," he says over morning tea in the society's august central-London headquarters.

Nurse wants the society to have a stronger voice on the big policy questions of the day. "The Royal Society has a responsibility to provide advice on difficult issues, even if they are contentious," he says.

He hopes to boost the society's role in government decision-making by fostering greater involvement of its roughly 1,500 fellows and foreign members in preparing reports,

potentially with the help of more policy staff. Nurse also wants to expand the number of authoritative and influential reports on key issues, such as nuclear power, climate change and the definition of life. The society has long produced such reports, most recently on the global scientific enterprise and on the potential threats and opportunities offered by geoengineering to mitigate climate change. But Nurse sees an opportunity to do more on a broader range of topics, with an eye for increasing the society's global reach. "I think the world would listen to us," he says.

Not everyone is convinced. "The first thing it should do is get a big bookshelf and put it in the basement to store the reports," says Daniel Greenberg, a journalist based in Washington DC who has devoted his career to studying the intersection of science and policy. The US National Academy of Sciences, which produces many more reports than its British counterpart, has relatively little influence

over the political process, he says. "Nobody in politics reads an academic report, slaps the side of their head and says 'Wow!'," he says.

Others suggest that the society could gain more influence by choosing its topics carefully. National Academy reports can sometimes shift the tone of a debate, says David Goldston, who was chief of staff on the US House Committee on Science from 2001 to 2006. And Robert May, a zoologist at the University of Oxford, UK, and former president of the society, points out that a 2002 report on foot-and-mouth disease helped to set up national vaccination strategies to prevent widespread cattle culling during future outbreaks. That report was successful in part because the society consulted closely with politicians and bureaucrats throughout, says May.

Nurse insists that the society will focus on "big areas that are important to our society" — not just those immediately relevant to policy-makers. Through conferences and studies, the society should also draw attention to the biggest mysteries in science, he says. "What is life? What is the beginning of the Universe? You know, that type of question." ■

➔ **NATURE.COM**
To read more about the history of the Royal Society, see: go.nature.com/yhtchtz



Coordinated research will help to reap the rewards of biofuel crops such as *Miscanthus × giganteus*.

BIOLOGY

US plant scientists seek united front

Academia and industry join forces to carve out ten-year plan.

BY HEIDI LEDFORD

The perennial grass *Miscanthus × giganteus* has all the makings of a biofuel superstar. It grows rapidly, converts sunlight into biomass ten times more efficiently than the average plant and has little need for fertilizer.

But *M. × giganteus* is a headache in the lab. Its genome has few markers to help would-be breeders keep track of desirable genes, and little is known about how it regulates important traits such as cold tolerance and water efficiency. It is also a sterile hybrid, complicating attempts at genetic improvement. "It has such great promise," says Neal Gutterson, president of Mendel Biotechnology, a company in Hayward, California, that is developing the grass as a biofuel crop. "But from a research perspective it is so painfully underdeveloped."

Gutterson hopes that the first ever summit to map the future of US plant science will change that, by encouraging researchers to tackle the genomic wilderness of emerging biofuel crops in a more systematic way.

The 22–23 September meeting, hosted by the Howard Hughes Medical Institute in Chevy Chase, Maryland, is the brainchild of Gary Stacey, an expert in host–microbe interactions in plants at the University of Missouri in Columbia. After a stint chairing the public-affairs committee for the American Society of Plant Biologists in Rockville, Maryland, Stacey says he realized that "we were speaking to Congress with too many dissonant voices. It was clear we had to start singing from the same hymnal."

At the summit, Stacey aims to bring together academic and industry scientists along with representatives from funding agencies and growers' associations to draw up a ten-year plan for plant biology. On the meeting's agenda are topics from bioenergy and informatics to the field's grand, overarching goal of predicting how a plant with a given set of genes will fare in different environments. The resultant list of priorities should aid coordination across a diverse research community and help to target the funds it receives from an array of federal sources. "It's a really smart

idea," says Karen Cone, a programme director at the US National Science Foundation (NSF). "It will give the plant-science community an opportunity to articulate a vision for the future that will influence the funding agencies."

Stacey began pushing for the meeting five years ago, borrowing a concept from US astronomers and astrophysicists, who survey their field once a decade to identify scientific priorities and rank potential projects. With the plant summit, he hopes to bridge a sometimes acrimonious divide between researchers who specialize in crops and those who work with model systems such as *Arabidopsis thaliana*, a quick-growing weed with a small genome that serves as a reference for plants that are harder to study. At the time, *Arabidopsis* researchers already had their own ten-year project: the *Arabidopsis* 2010 project funded by the NSF, which aimed to identify the function of every *Arabidopsis* gene by 2010.

Stacey says that the need for broad planning is now even greater because funding for the *Arabidopsis* 2010 project has run out and won't be extended (see *Nature* 464, 154; 2010). Many *Arabidopsis* researchers are now hoping to apply what they have learned from the weed to agriculturally important species with genomes once considered too big to tackle. "It is time to move forward into other species," says Cone.

As research objectives get more ambitious and cross species boundaries, plant scientists will need to coordinate their activities. "We're getting more like the physical sciences in the sense that we have to have bigger projects with enormous amounts of information," says Gutterson. "If you don't think about it ahead of time and create larger-scale interactions, you can't advance the science as effectively."

Although the decadal surveys of the astronomers take years to pull together, Stacey and organizers at the American Society of Plant Biologists hope to issue a report by early 2012, then circulate it to members of Congress and funding agencies.

"We were speaking with too many voices. It was clear we had to start singing from the same hymnal."

The team has already put together a list of about a dozen topics to be discussed over the two days and in wider solicitations to the community after the gathering.

A key to the summit's success, says Gutterson, will be the engagement of ecologists, whose expertise is becoming increasingly valuable because even molecular biologists are flocking to learn more about how the genes and processes they study function in natural environments. Stacey expects that summit participants, like all plant scientists, will tout their favourite species, but he hopes for unity in the programmes and technologies they push. "The tone might be different, but for once, we might be pulling together in the same direction." ■

HEALTH

UN targets top killers

International summit considers how to stem the rise in non-communicable diseases.

BY DECLAN BUTLER

When heads of state and health ministers gather in New York next week for the first United Nations (UN) high-level summit on non-communicable disease (NCD), they will be presented with some jaw-dropping statistics. According to UN reports released before the meeting, NCDs such as cardiovascular disease and cancer killed 36 million people in 2008, accounting for 63% of all deaths. Although NCDs are often mistakenly thought of as diseases of affluence, more than 80% of the NCD deaths occurred in low- and middle-income countries (see 'Total deaths'). By 2030, says the UN, the global annual toll of NCD will rise to 52 million deaths.

Total death statistics also suggest that apart from in the poorest countries in Africa, NCDs kill many more people than communicable diseases such as AIDS, malaria, tuberculosis or meningitis. This has led a growing number of health campaigners to demand global action on what they describe as an 'epidemic' of NCD. The summit has been promoted in particular by the NCD Alliance, an advocacy group launched in 2009 by four disease federations, including the World Heart Foundation in Geneva. The alliance campaigns for increasing support for research into NCD, strengthening health systems and reducing tobacco use, salt intake and other NCD risk factors.

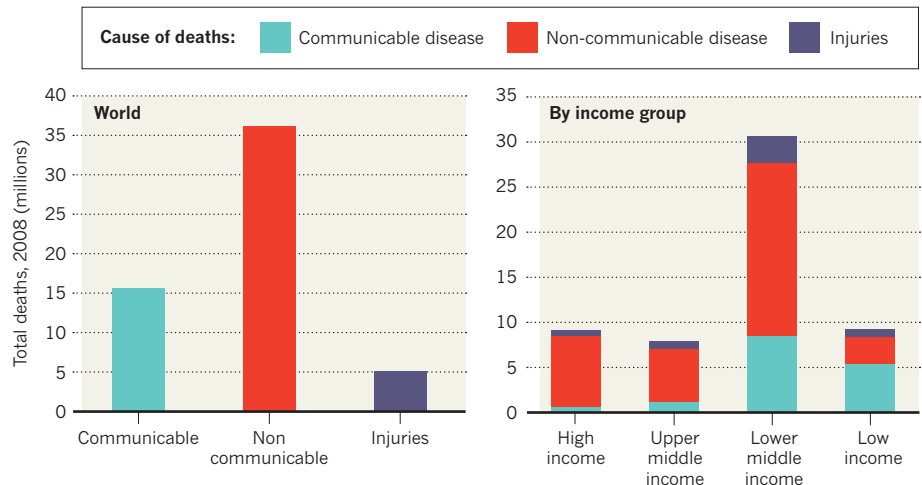
Ann Keeling, chair of the alliance and chief executive of the International Diabetes Federation in Brussels, says that whatever the outcome of the summit, the effort is "already a success" because it has put NCD high on the international political agenda.

But claims of an NCD epidemic could be missing a big part of the global picture. The predicted increases in total deaths are very real, but are not down to any sudden new disease risk, says Colin Mathers, coordinator of mortality and disease-burden statistics at the World Health Organization in Geneva. Almost all the extra deaths will stem from a current bulge in the number of young people in poorer countries, who will grow more susceptible to NCDs as they age. "It is not that the risk of disease for a given age is rising, but that there are more people," says Mathers.

And alarm over NCD can obscure the fact that infectious diseases still account for more years of life lost in many of the lower and middle-income countries,

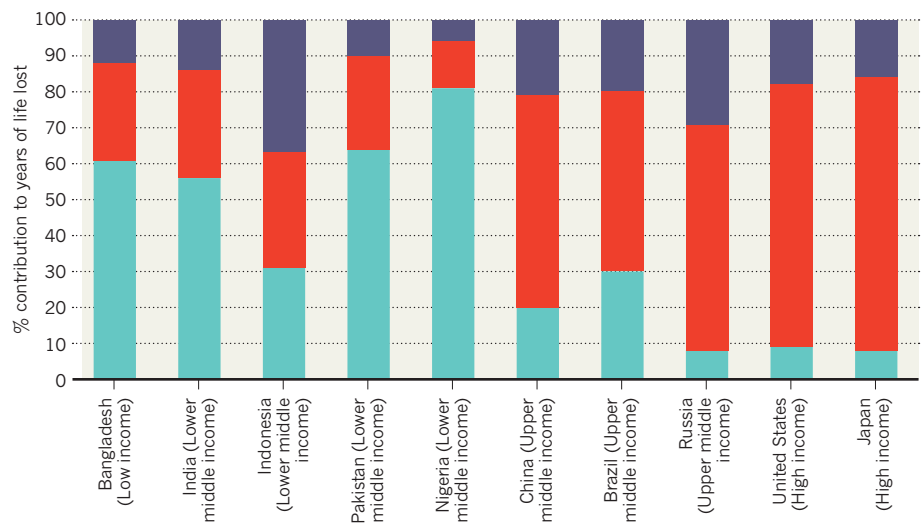
TOTAL DEATHS

Non-communicable disease (NCD) surpassed communicable disease as the greatest cause of all deaths in 2008, in all income groups except low-income countries. Middle-income countries have made progress against communicable disease in recent years, but a population bulge of young people, as well as increasing longevity, mean that more people have been exposed to NCD.



YEARS OF LIFE LOST

When mortality data are viewed in terms of 'years of life lost' rather than total deaths, the effect of NCD in lower-income countries becomes far less pronounced, with communicable diseases becoming more important. This chart shows the ten most-populous countries in 2004, accounting for about two-thirds of the world population.



because they often strike younger adults and children (see 'Years of life lost').

A lesser factor contributing to the rise in NCD deaths is that life expectancy in most low- and middle-income countries has risen spectacularly in recent decades, catching up with, and sometimes surpassing, those of higher income countries (see go.nature.com/idsgd1).

The probability of someone aged 15 dying before they are 60 — the '45q15' indicator — has likewise plummeted globally (see

'Premature death'). As a result, many more people are living long enough to develop NCD. Indeed, many indicators overlook the fact that global health is actually improving overall. Many Latin American countries that only a few decades ago were clawing their way out of poverty now have levels of health approaching those of Europe just 20 years ago, for example.

The exact trajectory that NCD will take in poorer countries over the next few decades is still an "open question", says Mathers. As

SOURCE: WHO

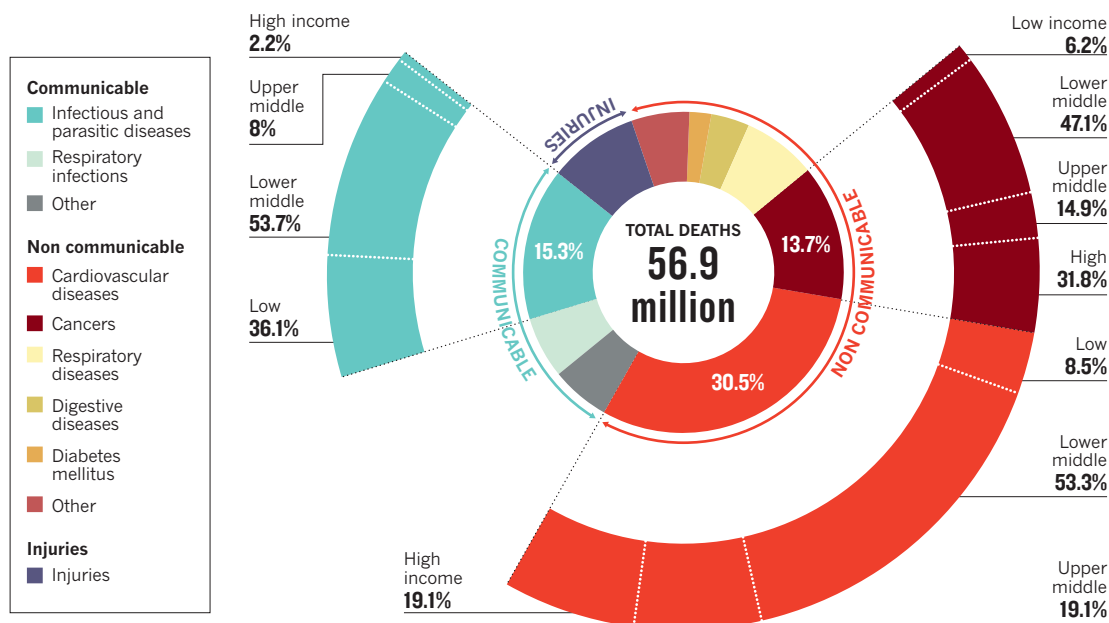
SOURCE: WHO

➔ **NATURE.COM**
For interactive
graphics, see:
go.nature.com/idsgd1

SOURCE: WHO

DEATHS BY CAUSE

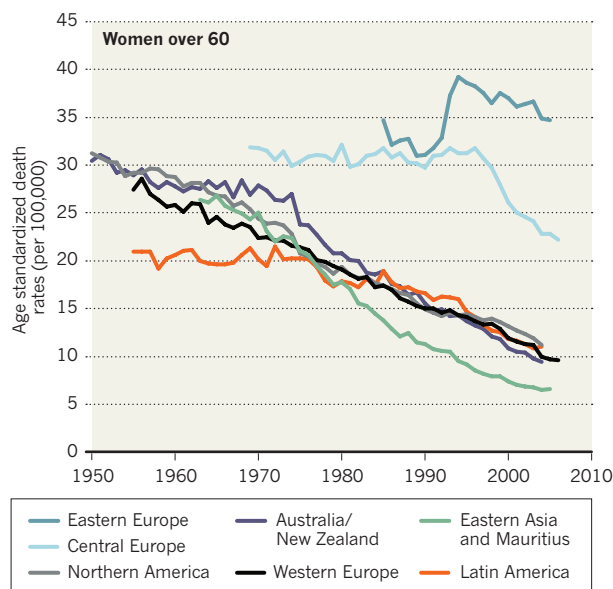
Cardiovascular diseases are the world's biggest killer, taking more than 17 million lives in 2008. As with cancer and infectious disease, lower-middle-income countries account for about half of the toll. But the remaining deaths from NCD are mostly in upper-middle- and high-income countries where treatments and preventative measures are already relatively well developed. By contrast, low-income countries — with the most room for health improvement — account for more than one-third of the toll for infectious diseases.



SOURCE: WHO

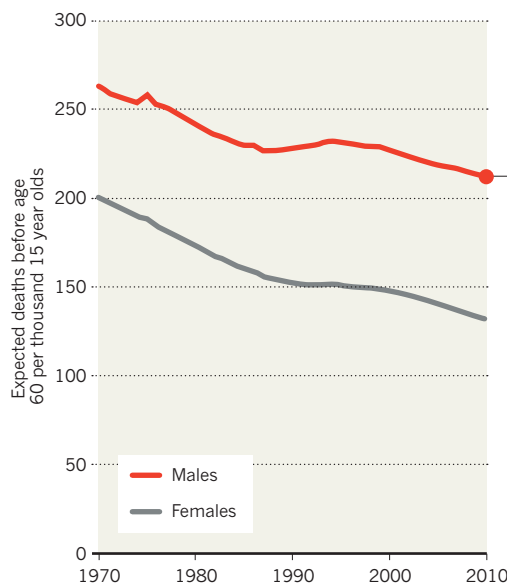
DECLINING CARDIOVASCULAR DEATH RATES

Many countries are successfully tackling NCD, with sharp falls in rates of cardiovascular disease, for example. Eastern Europe remains an outlier, with rapid growth in NCD following the collapse of the Soviet Union. Diabetes and women's lung cancer rates are also increasing globally.

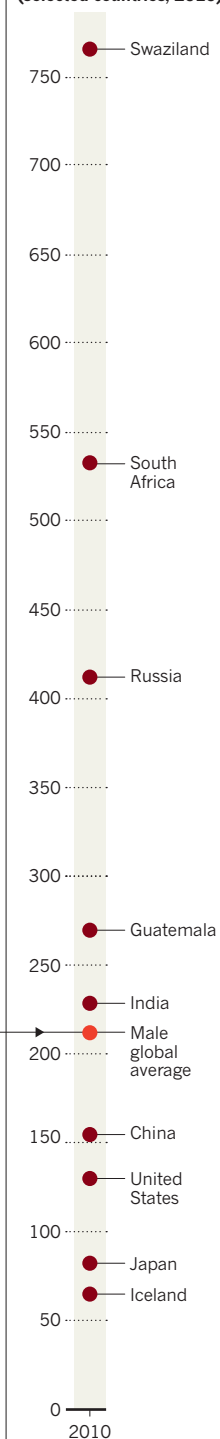


PREMATURE DEATH

The risk of adult mortality can be estimated using the '45q15' indicator — the proportion of 15 year-olds in any year who will die before reaching the age of 60. Despite large differences internationally (right), the global average is declining rapidly.



Male adult mortality risk (selected countries, 2010)



SOURCE: IHME

poorer countries grow wealthier, their health systems are likely to improve and drive down disease levels, for example. Per capita levels of many NCDs, including cardiovascular disease, have in fact fallen in most countries over the past few decades (see 'Declining cardiovascular death rates'). The most conspicuous exceptions are diabetes — on the rise because of increased obesity levels — and lung cancers in women, as a result of more women smoking.

The global trend in the rate of NCD

mortality, as opposed to absolute numbers of fatalities, is "down rather than up, especially in places where it is prioritized", concedes Johanna Ralston, chief executive of the World Heart Federation. "Where it is not yet on the agenda, it doesn't get prioritized, which is why [the UN summit] is so important."

The case for action in poorer countries is compelling, she adds. The rate of increase of total NCD deaths in poorer countries is faster than in the past, potentially overwhelming

underdeveloped health systems, she says.

Indeed, the outlook for poorer countries that have dysfunctional health-care systems, or that fail to tackle disease risk factors, could be bleak, says Mathers. In West Africa, for example, high blood pressure is common yet often goes untreated, even though cheap drugs are available. "There are enormous amounts of NCD in low-income countries that are preventable, but which aren't being prevented because of failed health systems," he says. ■ SEE EDITORIAL P.250

ENVIRONMENT

Science enters desert debate

United Nations considers creating advisory panel on land degradation akin to IPCC.

BY NATASHA GILBERT

A desert may need no defining, but desertification is not so easy to pin down. Although the loss of soil nutrients and moisture threatens roughly a third of the world's land area, imperilling farming and biodiversity, scientists lack a clear definition of it or agreed standards to measure its causes and progression. That absence has hampered global efforts to tackle these problems under the United Nations Convention to Combat Desertification (UNCCD) — unable to track the impact of their funding, donors are reluctant to invest.

Next week, a meeting of UN member states in New York will begin to set out the steps needed to close that science gap and boost international efforts to tackle desertification. Possible actions include improving irrigation, increasing fertilization and promoting the use of land-management techniques that integrate forests with agriculture. The gathering will set the tone for a broader meeting, to be held on 10–21 October in Changwon, South Korea, of the 193 nations that are party to the UNCCD.

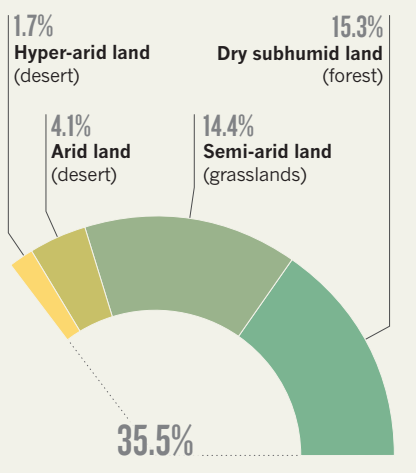
SYMPTOMS AND CAUSES

The UNCCD has had limited success since it came into force in 1996. “Donors must have a clear idea of how big the problem is and must be confident that we can measure progress in overcoming the problem,” says William Dar, director-general of the International Crops Research Institute for the Semi-Arid Tropics in Andhra Pradesh, India.

Monitoring and assessment have so far focused mainly on the symptoms of land degradation and desertification, such as loss of top soil and decreased food production. But researchers need to understand how societal and economic factors, including poverty and child malnutrition, can drive these processes. They also need to learn how best to track desertification using satellite data. Remote

LIVING ON DRY EARTH

Roughly one-third of the world's population lives in dry-land areas, according to the most recent data from the UN's Millennium Ecosystem Assessment.



sensing can measure surface temperature and vegetation cover, for example, but those data can reflect temporary heat waves and dry spells rather than long-term desertification. As a first step, in 2012, nations that are party to the convention will provide data on two measurable indicators: the proportion of the population in vulnerable areas living above the poverty line and the area of land covered by vegetation. This will begin to provide a baseline from which to measure if and how land is degrading.

Governments of some developed nations also argue that land degradation is not a global issue of immediate concern to their citizens, but rather a problem limited to the dry lands of developing countries. Indeed, the convention applies only to lands that are already classed as ‘dry’, including arid and semi-arid regions, along with the dry subhumid (mainly forested) areas of Australia, China and Russia (see ‘Living on dry earth’).

The convention's secretariat now wants to broaden its scope to include humid and wet lands that are at risk of future degradation through the effects of climate change, for example. “Desertification can occur anywhere except in a desert,” says Uriel Safriel, a desert ecologist at the Hebrew University of Jerusalem, Israel. This move could make the convention more directly relevant to rich nations.

“This is a turning point for the convention,” says Mansour N'Diaye, acting deputy executive secretary to the convention. Next week's meeting is expected to propose the establishment of a scientific advisory panel on land and soil degradation, akin to the Intergovernmental Panel on Climate Change, which advises the convention's sister body on climate change. The panel would review the latest scientific data on the extent of land degradation, propose and assess efforts to combat it, and perhaps also develop specific targets for halting degradation, says Dar. But he notes that the panel's success would depend on ensuring that it was free from political interference and had the scientific independence needed to provide unbiased advice.

Safriel supports the secretariat's push for greater involvement of developed nations by highlighting the global nature of the problem. But he worries that developing countries may object, as it could draw attention away from

“It's stupid to stick to dry lands. It's a global issue.”

desertification in their own lands. Likewise, developed countries may be reluctant to commit funds and resources, arguing that existing initiatives, such as the UN's Millennium Development Goals to improve the lot of the world's poorest people, already address these issues. Despite these concerns, Safriel says, it is vital to broaden the convention's focus. “It's stupid to stick to dry lands, because not only dry lands will become dry. It's a global issue.” ■



**MORE
ONLINE**

TOP STORY



Germany learns from its latest *E. coli* outbreak
go.nature.com/gj9pqc

MORE NEWS

- Seals slide towards extinction in Hawaiian reserve go.nature.com/s9mntl
- Miniature microscopes capture neurons in action go.nature.com/zdblija
- Detector suggests dark-matter particles may be lighter than expected go.nature.com/9uyptz

SLIDESHOW



Jupiter shot takes prize for heavenly images
go.nature.com/3khu94



S. SCHAEFFER

Q&A Subra Suresh

Merit comes first

Subra Suresh, former dean of engineering at the Massachusetts Institute of Technology in Cambridge, began a six-year term as the director of the National Science Foundation (NSF) in Washington DC last October. As he nears his one-year anniversary as head of the US\$6-billion agency — the primary funder of basic physical-sciences research in the United States — Suresh discusses the challenges that the NSF faces, including a stormy fiscal climate and mounting calls for research to show economic returns.

How important is it that science looks useful in the current budget climate?

It's important for us to articulate the usefulness of science. But for an agency like the NSF, it will be dangerous if we choose to follow the latest fashion and lose sight of the long-term need to support science. We are not a mission agency. We don't have to produce a product next year.

In July, the NSF launched the Innovation Corps (I-Corps) to support commercialization of technology. Isn't that supposed to help increase pay-offs in the short term?

I-Corps' goal is to create a national infrastructure and mechanisms that help to support those institutions that have already received funding from the NSF.

➔ **NATURE.COM**
For commentaries on the Innovation Corps and merit review, see: go.nature.com/mzyqmh
go.nature.com/lzlvct

Suppose that a young inventor, supported by the NSF, needs to know whether a scientific

discovery has any potential value beyond a publication. Who do they talk to? Where do they go? Usually, at larger universities, you talk to the technology-licensing officer. But there are many institutions where the young faculty member doesn't have that opportunity.

So would a measure of the programme's success be an improvement in such an institution's patent rate?

It could be patent rate, or patents that are licensed, not just patents that are filed. It could also be how industry has engaged with that university, and helped to educate the students.

The National Science Board, the NSF's governing body, has a task force that is evaluating the two funding criteria: intellectual merit and broader impacts. How would you define broader impacts?
Broader impact has many different flavours. In some cases, the science itself can have a broad

societal impact. You could do something with NSF-funded research that changes the way undergraduate education happens around the country: create a tool, a mechanism, a model, software. That's broader impact.

Bringing under-represented groups to your lab, or giving talks in community colleges about your research, exciting those who would otherwise not go into science. That's also broader impact.

How much should researchers be thinking about broader impacts?

We don't want to compromise scientific merit at all; the work has to be outstanding before it gets funding. But given that, I think it's useful to ask the question: what's the impact of this work? That is why, in 1997, the NSF created broader impact as an additional criterion.

Whose responsibility is it to ensure that work has a broader impact?

Individuals can play an important part, as can institutions and departments. This is a delicate balance to achieve. You cannot say that individuals are responsible and the institution is not, or vice versa.

How low can the grant acceptance rate be and still be acceptable?

There is no magic number. Low success rates can be devastating for a number of reasons: you have to write more proposals before you're successful; it can damage morale; it strains the reviewer system, which is our peer community.

What are the mechanisms we can put in place so that we don't waste the community's time? The National Institutes of Health has a triage system. Is that the right system for the NSF? Maybe, maybe not. In some cases, we have preproposals — shorter proposals. Can that be a mechanism? Can we put a cap on how many NSF proposals you can submit at a given time? Right now, there's no upper limit.

Is that something you're considering?

It's under discussion. Everything is under discussion. But at the same time, we want to be fair. We want to be sure that people who are very good, who should be in science, don't get discouraged because two of their proposals didn't get funded.

You're still quite new to Washington. What has surprised you most?

I have to mention a quote I've heard attributed to Woodrow Wilson, who was president of Princeton University in New Jersey before he was president of the United States. When he arrived in Washington, a reporter asked him why he left his Ivy League school. Apparently, Mr Wilson responded: "I came to Washington so I don't have to deal with politics anymore." Some of this is new, but not all of it. ■

INTERVIEW BY ERIC HAND



AT FAULT?

In 2009, an earthquake devastated the Italian city of L'Aquila and killed more than 300 people. Now, scientists are on trial for manslaughter.

BY STEPHEN S. HALL IN L'AQUILA, ITALY



A. NUSCA/POLARIS/EYEVINE

was no imminent danger, and recalling scientific statements claiming that each shock diminished the potential for a major earthquake, he persuaded his family to remain in their apartment on Via Luigi Sturzo. All three of them were huddled together in the master bed when, at 3:32 a.m. on 6 April, a devastating magnitude-6.3 earthquake struck the city.

"It was like being in a blender," Vittorini recalls. "It wasn't a roar, it was a gigantic noise. And then darkness." The apartment building, a structure of reinforced concrete constructed in 1962, instantly collapsed, and their third-floor apartment ended up in a jumble of wreckage several feet off the ground. Seven people were killed in the collapse of the building, including Vittorini's wife and daughter; he was pulled from the rubble, injured but alive, six hours later. The earthquake claimed 309 lives in L'Aquila and several towns nearby, injured more than 1,500 people, destroyed some 20,000 buildings and left 65,000 people temporarily displaced.

The apartment building on Via Luigi Sturzo is "just a hole now," Vittorini says, and his childhood home and the piazza where families spent the night are, like almost all of L'Aquila's historic centre, now in a barricaded and inaccessible 'red zone'. More than two years after the earthquake, block after block of elegant, centuries-old buildings is corseted by bands of structural reinforcement; wooden braces prop up numerous Gothic windows and arches in uninhabitable buildings. The basilica of San Bernardino, the city hall, the Cinema Massimo — all closed. On a cracked ochre wall along the main corso, one of the few streets that remain open in the centre, someone has scribbled in black paint: "*L'Aquila é morta*." (L'Aquila is dead.)

In a trial set to begin next week, an Italian judge will decide whether the symbolic death of L'Aquila — and, more specifically, the earthquake-related deaths of dozens of citizens included in the lawsuit, including Vittorini's wife and daughter — constituted a crime due to the negligence of six leading Italian scientists and one government official, who have been charged with manslaughter in connection with the case.

When the charges were first aired in June 2010 by public prosecutor Fabio Picuti, the case was likened to a frivolous attempt by overzealous local prosecutors to make scapegoats out of some of Italy's most respected geophysicists: Enzo Boschi, then-president of Italy's National Institute of Geophysics and Volcanology (INGV) in Rome; Franco Barberi, at the University of Rome Tre; Mauro Dolce, head of the seismic-risk office at the national Department of Civil Protection in Rome; Claudio Eva, from the University of Genova; Giulio Selvaggi, director of the INGV's National Earthquake Centre in Rome; and Gian Michele Calvi, president of the European Centre for Training and Research in Earthquake Engineering in Pavia; as well as government official Bernardo De Bernardinis, then vice-director of the Department of Civil Protection. According to an open letter to the president of Italy, Giorgio Napolitano, signed by more than 5,000 members of the scientific community, the seven Italians essentially face criminal charges for failing to predict the earthquake — even though pinpointing the time, location and strength of a future earthquake in the short term remains, by scientific consensus, technically impossible.

The indictments have drawn global condemnation. The American Geophysical Union and the American Association for the Advancement of Science (AAAS), both in Washington DC, issued statements in support of the Italian defendants. In an open letter to Napolitano, for example, the AAAS said it was "unfair and naive" of local prosecutors to charge the men for failing "to alert the population of L'Aquila of an impending earthquake". And last May, when Italian magistrate Giuseppe Gargarella ruled at a preliminary hearing that the scientists would have to stand trial this September, the Italian blogosphere lit up with lamentation and defence lawyers greeted the decision with disbelief. "On the one hand, he's stunned," Francesco Petrelli said of his client, Barberi. "On the other, he's very pained and sad."

The view from L'Aquila, however, is quite different. Prosecutors and the families of victims alike say that the trial has nothing to do with

From when he was a young boy growing up in a house on Via Antinori in the medieval heart of this earthquake-prone Italian city, Vincenzo Vittorini remembers the ritual whenever the family felt a seismic tremor overnight. "My father was afraid of earthquakes, so whenever the ground shook, even a little, he would gather us and take us out of the house," he says. "We would walk to a little piazza nearby, and the children — we were four brothers — and my mother would sleep in the car. My father would stand outside, smoking cigarettes with the other fathers, until morning." That, he says, represented the age-old, cautionary "culture" of living in an earthquake zone.

Vittorini, a 48-year-old surgeon who has lived in L'Aquila all his life, will never forgive himself for breaking with that tradition on the night of 5 April 2009. After hundreds of low-level tremors over several months, L'Aquila shook with a strong, magnitude-3.9 tremor shortly before 11 p.m. on that Palm Sunday evening. Vittorini debated with his wife Claudia and his terrified nine-year-old daughter Fabrizia whether to spend the rest of the night outside. Swayed by what he describes as "anaesthetizing" public assurances by government officials that there



IGNACIO MARIA COCCIA/LUZPHOTO/ EYEVINE



Vincenzo Vittorini's apartment building collapsed in the 2009 quake, killing his wife and daughter. He says that he feels "betrayed by science".

IGNACIO MARIA COCCIA/LUZPHOTO/ EYEVINE

the ability to predict earthquakes, and everything to do with the failure of government-appointed scientists serving on an advisory panel to adequately evaluate, and then communicate, the potential risk to the local population. The charges, detailed in a 224-page document filed by Picuti, allege that members of the National Commission for Forecasting and Predicting Great Risks, who held a special meeting in L'Aquila the week before the earthquake, provided "incomplete, imprecise, and contradictory information" to a public that had been unnerved by months of persistent, low-level tremors. Picuti says that the commission was more interested in pacifying the local population than in giving clear advice about earthquake preparedness.

"I'm not crazy," Picuti says. "I know they can't predict earthquakes. The basis of the charges is not that they didn't predict the earthquake. As functionaries of the state, they had certain duties imposed by law: to evaluate and characterize the risks that were present in L'Aquila." Part of that risk assessment, he says, should have included the density of the urban population and the known fragility of many ancient buildings in the city centre. "They were obligated to evaluate the degree of risk given all these factors," he says, "and they did not."

"This isn't a trial against science," insists Vittorini, who is a civil party

to the suit. But he says that a persistent message from authorities of "Be calm, don't worry", and a lack of specific advice, deprived him and others of an opportunity to make an informed decision about what to do on the night of the earthquake. "That's why I feel betrayed by science," he says. "Either they didn't know certain things, which is a problem, or they didn't know how to communicate what they did know, which is also a problem."

Although the outcome of the trial may not be known for months, if not years, the events leading up to the earthquake are already being viewed as a sobering case study in risk assessment and public communication — a scenario that might easily be replayed in a future that includes not just 'conventional' natural disasters (such as volcanic eruptions, earthquakes, and tsunamis), but also extreme weather events (such as tornadoes, hurricanes, floods and droughts) perhaps cooked up by climate change. The trial has already had a chilling effect on scientists' willingness to share their expertise with the public. "When people, when journalists, asked my opinion about things, I used to tell them, but no more. Scientists have to shut up," says Boschi, whose successor at the INGV was appointed last month. Others see the case as an indictment of the obfuscating, probabilistic language with which scientists characterize the uncertain potential of natural disasters. Selvaggi, one of the indicted scientists, says that the charges serve as a "dangerous" warning to researchers, who may find themselves in legal trouble because of the way that non-scientists such as public officials or journalists translate their risk analyses for public consumption. Given the novelty of the issues, says defence lawyer Filippo Dinacci, "not only the press, but the academic legal community will be watching this case with great interest".

Thomas Jordan, director of the Southern California Earthquake Center at the University of Southern California in Los Angeles, and chair of the International Commission on Earthquake Forecasting (ICEF), which reviewed the L'Aquila events in a report released in May,

says that in his view the prosecution charges have “no merit”. But he adds that the trial is nonetheless a “watershed case” that will force seismologists worldwide to rethink the way they describe low probability, high-risk events, as well as an opportunity for the scientific community at large to assess “rising public expectations” about how information on natural disasters should be handled. “The public expects authoritative, transparently available information,” he says, “and we need to say what we know in an explicit way.”

In Jordan's view, “It has to be done right, and it was not in L'Aquila.”

SEISMIC REPUTATION

L'Aquila is — or was — a jewel of medieval beauty set in the middle of one of the most seismically dangerous zones in Italy. Surrounded by the massive peaks of the restless Apennine mountain range, the city, capital of the Abruzzo region, was largely destroyed by earthquakes in 1461 and in 1703. Its seismic reputation was such that the nineteenth-century British travel writer Augustus Hare noted that, “nature suddenly often sets all the bells ringing and the clocks striking, and makes fresh chasms in the old yellow walls”.

Its most recent seismic tragedy began in October 2008, when dozens of low-magnitude tremors began to hit the city and surrounding areas along the Aterno River valley (see ‘A shaken city’). Known as seismic swarms, these tremors continued intermittently over the first three months of 2009; according to Picuti, they numbered 69 in January, 78 in February and 100 in March, with an additional 57 shocks during the first five days of April. “It was like this almost every day,” says Pier Paolo Visione, a local accountant, shaking a table in a restaurant with a slow but vigorous motion that nearly topples a bottle of the local red Montepulciano wine. “I had never been afraid of earthquakes before, but my skin began to crawl.” (Visione's sister died in the quake, and he is a civil party to the suit.)

Unnerving though these clusters may be, experts agree that seismic swarms rarely precede major earthquakes. In 1988, seismic engineer Giuseppe Grandori, now professor emeritus at the Polytechnic of Milan, and his colleagues published a retrospective analysis of seismic swarms in three other earthquake-prone Italian localities (G. Grandori *et al. Bull. Seismol. Soc. Am.* **78**, 1538–1549; 1988). They concluded that a medium-sized shock in a swarm forecasts a major event within several days about 2% of the time, and Grandori says that the same was probably true for the region around L'Aquila.

Translating these risks is extremely challenging for civil defence officials. In Grandori's view, there is a 98% probability of a false alarm if officials issue an alert, yet a terrible price to pay in loss of life and property if they fail to issue a warning and a major quake occurs. After a medium-sized shock in a seismic swarm, the risk of a major quake can increase anywhere from 100-fold to nearly 1,000-fold in the short term, according to Jordan, although the overall probability remains extremely low. “What do you tell people in that situation?” he says. “You're sort of between Scylla and Charybdis on this thing.”

To this difficult exercise in risk probability was added a wild card in the case of L'Aquila: a resident named Giampaolo Giuliani began to make unofficial earthquake predictions on the basis of measurements of radon gas levels. Giuliani, who had worked for 40 years as a laboratory technician, including 20 years at the nearby Gran Sasso National Laboratory until his retirement in 2010, had deployed four home-made radon detectors throughout the region.

The idea behind radon measurement, Giuliani says, is that emissions of the gas fluctuate significantly in the 24 hours before an earthquake. But their use as a reliable short-term predictor of earthquakes has never been scientifically proved or accepted. The recent ICEF report deemed Giuliani's findings “unsatisfactory”, and he has yet to publish a single

peer-reviewed paper on his radon work. Nonetheless, he maintained an open website that posted real-time radon measurements from his detectors, and in interviews with journalists and in an informal mobile-phone network, Giuliani made predictions about low-level seismic activity. Although the ICEF report notes that he made two false forecasts, *The Guardian* newspaper dubbed him “The Man Who Predicted An Earthquake”, after the April 2009 quake hit.

As word spread about Giuliani's unofficial predictions, even more unease percolated through the population. Marcello Melandri, the lawyer for Boschi, says that Giuliani had been terrifying local residents, and that Guido Bertolaso, head of Italy's Department of Civil Protection agency, “was very worried about the population of L'Aquila”. On 30 March, Giuliani says, national civil-protection officials cited him

for *procurato allarme* — essentially instigating public alarm or panic — and forbade him from making any public pronouncements.

That same day, L'Aquila was hit by an intense, magnitude 4.1 shock in the afternoon that deeply rattled local residents. Vittorini, who performs his surgeries in the nearby town of Popoli, received an anguished call from his wife and son. (His daughter was not at home at the time.) He urged them to leave the house immediately and get outside, he says. L'Aquila's mayor, Massimo Cialente, ordered the evacuation of several public buildings and closed the De Amicis primary school to inspect for structural damage.

Italian seismologists had been monitoring the swarm in the Abruzzo region for months, and notifying civil-protection officials in real time of every tremor with a magnitude of greater than 2.5. Now, given the growing unease in L'Aquila, Bertolaso decided to convene an unusual meeting of the risks commission. The commission normally meets in Rome to assess the probability of earthquakes, volcanoes and other natural disasters, but this meeting was to take place the next day in L'Aquila. The goal, according to a press release from the Department of Civil Protection, was to furnish citizens in the Abruzzo region “with all the information available to the scientific community about the seismic activity of recent weeks”.

MEETING OF MINDS

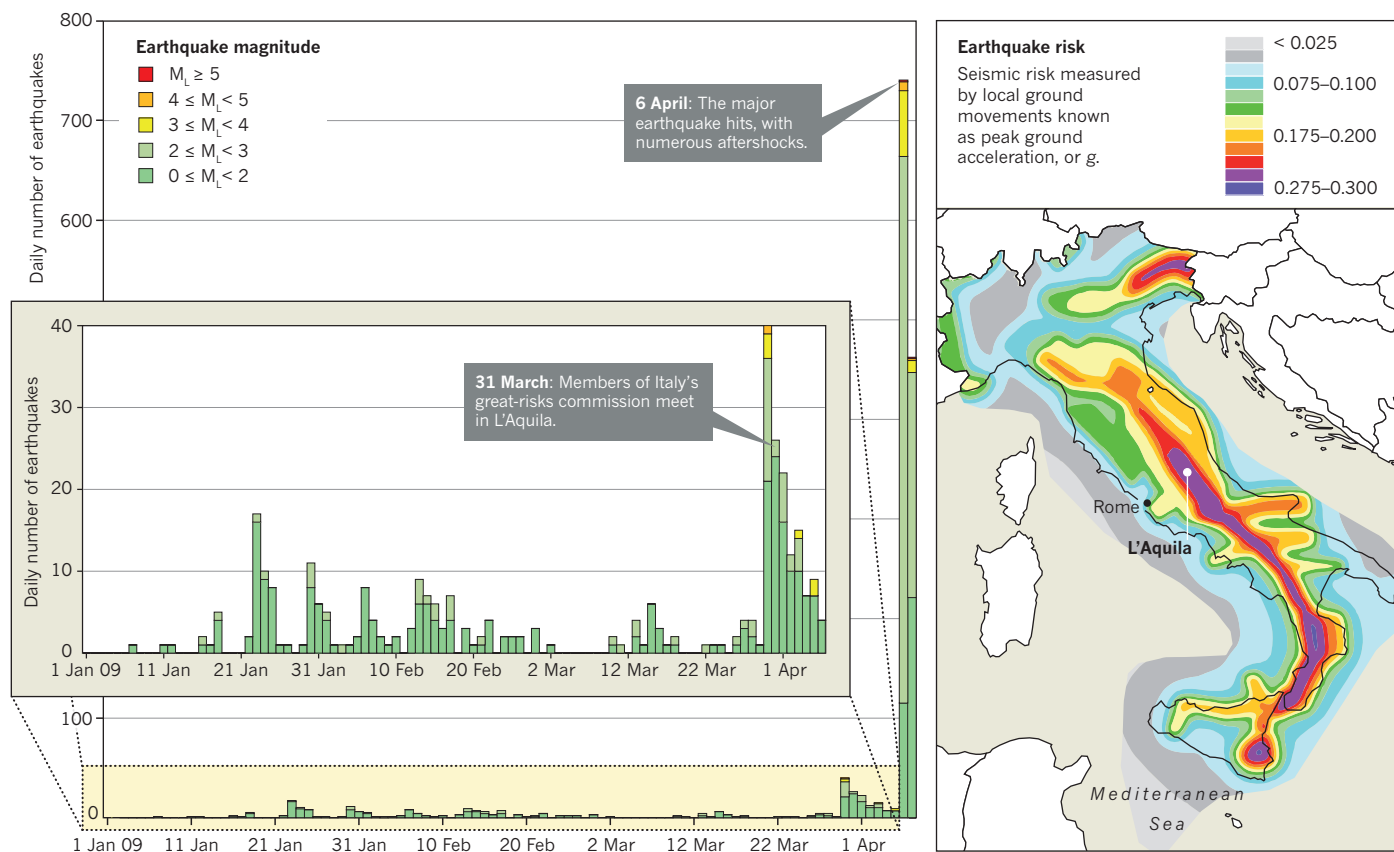
The now-famous commission meeting convened on the evening of 31 March in a local government office in L'Aquila. Boschi, who had travelled by car to the city with two other scientists, later called the circumstances “completely out of the ordinary”. Commission sessions are usually closed, so Boschi was surprised to see nearly a dozen local government officials and other non-scientists attending the brief, one-hour meeting, in which the six scientists assessed the swarms of tremors that had rattled the local population. When asked during the meeting if the current seismic swarm could be a precursor to a major quake like the one that levelled L'Aquila in 1703, Boschi said, according to the meeting minutes: “It is unlikely that an earthquake like the one in 1703 could occur in the short term, but the possibility cannot be totally excluded.” The scientific message conveyed at the meeting was anything but reassuring, according to Selvaggi. “If you live in L'Aquila, even if there's no swarm,” he says, “you can never say, ‘No problem.’ You can never say that in a high-risk region.” But there was minimal discussion of the vulnerability of local buildings, say prosecutors, or of what specific advice should be given to residents about what to do in the event of a major quake. Boschi himself, in a 2009 letter to civil-protection officials published in the Italian weekly news magazine *L'Espresso*, said: “actions to be undertaken were not even minimally discussed”.

Many people in L'Aquila now view the meeting as essentially a public-relations event held to discredit the idea of reliable earthquake prediction (and, by implication, Giuliani) and thereby reassure local residents. Christian Del Pinto, a seismologist with the civil-protection

“EITHER THEY DIDN'T KNOW CERTAIN THINGS, WHICH IS A PROBLEM, OR THEY DIDN'T KNOW HOW TO COMMUNICATE WHAT THEY DID KNOW, WHICH IS ALSO A PROBLEM.”

A SHAKEN CITY

L'Aquila lies in one of the most seismically hazardous zones in Italy (see map). In early 2009, a series of tremors hit the region. The graph shows the daily number of earthquakes, with the bar colour indicating magnitude. The tremors (shown in detail in the inset graph) were followed on 6 April by a devastating magnitude-6.3 earthquake, which killed more than 300 people.



SOURCE: INGV

department for the neighbouring region of Molise, sat in on part of the meeting and later told prosecutors in L'Aquila that the commission proceedings struck him as a "grotesque pantomime". Even Boschi now says that "the point of the meeting was to calm the population. We [scientists] didn't understand that until later on."

What happened outside the meeting room may haunt the scientists, and perhaps the world of risk assessment, for many years. Two members of the commission, Barberi and De Bernardinis, along with mayor Cialente and an official from Abruzzo's civil-protection department, held a press conference to discuss the findings of the meeting. In press interviews before and after the meeting that were broadcast on Italian television, immortalized on YouTube and form detailed parts of the prosecution case, De Bernardinis said that the seismic situation in L'Aquila was "certainly normal" and posed "no danger", adding that "the scientific community continues to assure me that, to the contrary, it's a favourable situation because of the continuous discharge of energy". When prompted by a journalist who said, "So we should have a nice glass of wine," De Bernardinis replied "Absolutely", and urged locals to have a glass of Montepulciano.

The suggestion that repeated tremors were favourable because they 'unload', or discharge, seismic stress and reduce the probability of a major quake seems to be scientifically incorrect. Two of the committee members — Selvaggi and Eva — later told prosecutors that they "strongly dissented" from such an assertion, and Jordan later characterized it as "not a correct view of things". (De Bernardinis declined a request for an interview through his lawyer, Dinacci, who insisted that De Bernardinis's public comments reflected only what the commission scientists had told him. There is no mention of the discharge idea in the official minutes, Picuti says, and several of the indicted scientists point out that De

Bernardinis made these remarks before the actual meeting.)

That message, whatever its source, seems to have resonated deeply with the local population. "You could almost hear a sigh of relief go through the town," says Simona Giannangeli, a lawyer who represents some of the families of the eight University of L'Aquila students who died when a dormitory collapsed. "It was repeated almost like a mantra: the more tremors, the less danger." "That phrase," in the opinion of one L'Aquila resident, "was deadly for a lot of people here."

The press conference and interviews, prosecutors argue, carried special weight because they were the only public comments to emerge immediately after the meeting. The commission did not issue its usual formal statement, and the minutes of the meeting were not even prepared, says Boschi, until after the earthquake had occurred. Moreover, it did not issue any specific recommendations for community preparedness, according to Picuti, thereby failing in its legal obligation "to avoid death, injury and damage, or at least to minimize them".

Picuti argues that the fragility of local housing should have been a central component in the commission's risk assessment. "This isn't Tokyo, where the buildings are anti-seismic," he says. "This is a medieval city, and that raises the risk." In 1999, Barberi himself had compiled a massive census of every seismically vulnerable public building in southern Italy; the survey, according to the prosecution brief, indicated that more than 550 masonry buildings in L'Aquila were at medium-high risk of collapsing in the event of a major earthquake.

The failure to remind residents of earthquake preparedness procedures in the face of such risks is one of the reasons that John Mutter, a seismologist at Columbia University's Lamont-Doherty Earth Observatory, declined to sign the open letter circulated to support the Italian scientists. Mutter says that in his opinion, "these guys shouldn't go to

jail, but they should be fined or censured because they should have said something other than what they said. To say 'don't worry' — that sort of thing just isn't helpful. You need to remind people of their earthquake drills: if they feel the house moving, get out of the building if you can, or get under a table or a door frame if you can't. Do all the things that we know save lives."

As part of the prosecution's case, Picuti argues in his brief that local residents made fateful decisions on the night of the earthquake on the basis of statements made by public officials outside the meeting. Maurizio Cora, a lawyer who lived not far from Vittorini, told prosecutors that after the 30 March shock, he and his family retreated to the grounds of L'Aquila's sixteenth-century castle; after the 11 p.m. foreshock on 5 April, he said his family "rationally" discussed the situation and, recalling the reassurances of government officials that the tremors would not exceed those already experienced, decided to remain at home, "changing our usual habit of leaving the house when we felt a shock". Cora's wife and two daughters died when their house collapsed.

"That night, all the old people in L'Aquila, after the first shock, went outside and stayed outside for the rest of the night," Vittorini says. "Those of us who are used to using the Internet, television, science — we stayed inside."

DISPUTED ADVICE

In an interview in the Rome offices of his lawyer, Boschi derided as "absurd" the idea that he in any way played down the risk to L'Aquila. Brandishing a copy of the INGV's seismic hazard map of Italy, which shows a broad swath of the Apennines in bright hues indicating high risk, the tall, silver-haired geophysicist insisted: "No one can find a single piece of paper where I say, 'Be calm, don't worry'. I have said for years that the Abruzzo is the most seismologically dangerous zone in all of Italy. It's as if I suddenly became an imbecile. I'm accused of being negligent!" He was not invited to participate in the press conference after the meeting, he says, and didn't even know about it until after his return to Rome.

Attorneys for the other scientists all insist that the charges are without foundation, while raising additional arguments. Barberi's lawyer, Petrelli, acknowledges that the meeting was intended "in part" to defuse the panic over Giuliani's predictions, but insists that everything his client said was scientifically sound and correct. To convey the difficulty of communicating risk assessments, he offers the analogy of being asked the safest way to travel, and recommending flying because it is statistically much safer than car or train. "If the person takes the plane, and the plane is involved in an accident, this doesn't mean that my advice was wrong," he said. "I gave the right advice, since scientific advice is based on statistics, and the statistics don't exclude the possibility of an event that we would like to avoid."

Alessandra Stefano, the lawyer for Calvi, says that the mass media has played a part in the case by disseminating incorrect information about "especially delicate" scientific matters. Eva's lawyer, Alfredo Biondi, has pointed out that in 1985, the then-head of civil protection, Giuseppe Zamberletti, was investigated for instigating a public panic when he ordered the evacuation of several villages in northwest Tuscany after a seismic swarm; on that occasion, no major quake occurred. Antonio Pallotta has argued that his client, Selvaggi, was not an official member of the commission.

As for the statement that seems to have resonated most with the residents of L'Aquila — De Bernardinis's claim that during seismic swarms, repeated tremors were "favourable" — Dinacci says of his client: "He's not a seismologist, he's a hydraulic engineer," and that he had only relayed what the scientists had told him. As to De Bernardinis's suggestion to have a glass of Montepulciano, Dinacci says, "This was a joke!

**"TO HAVE MADE A JOKE
ABOUT A GLASS OF
WINE AND THEN FACE A
CONVICTION IS ABSURD.
IT'S SOMETHING OUT OF
THE MIDDLE AGES."**

To have made a joke about a glass of wine and then face a conviction is absurd. It's something out of the Middle Ages."

The outcome of the trial that begins next week in L'Aquila can no more be predicted than can earthquakes themselves. It will ultimately be up to a single magistrate to decide whether the actions of the commission, and the alleged "erroneous information" released by officials outside the meeting, rise to the level of criminal culpability. Although defence lawyers say that the prosecution's case is logically flawed, the stakes are high. If convicted, the scientists could face up to 15 years in jail, according to prosecutors. In addition, plaintiffs in a separate civil case are seeking damages in the order of €22.5 million (US\$31.6 million).

AFTER SHOCK

Irrespective of the verdict, the episode has been a painful tutorial about the importance of clear public communication when potential disasters loom. The commission and the civil-protection department "got trapped in the wrong conversation because of the hullabaloo that was happening" around the unofficial predictions of earthquakes, says Jordan. "The issue became, is there going to be an earthquake or not, and that choice is the wrong way to talk about this." Mutter adds that in his opinion,

the commission's focus on whether earthquakes could be predicted or not ultimately didn't tell people what they wanted to know. "People aren't stupid," he says. "They know we can't predict earthquakes. They just want clear advice on what they should do."

The recent ICEF report argues that frequently updated hazard probabilities are the best way to communicate risk information to the public. "Seismic weather reports, if you will, should be put out on a daily basis," Jordan says. "Nobody has set up a good system for doing this, and our understanding of the 'weather' in this case is very poor, so we can only see through the glass darkly." But in an age of social media and instantaneous communication, he says, misinformation travels fast, and the public needs clear, real-time risk assessment. As Selvaggi warns, the number of situations in which scientists are asked to assess hazard is certain to rise. "We have an increasing number of extreme events," he said, "and we have increasing numbers of people living in high-risk regions. It's time to address this problem."

Jordan says that the L'Aquila incident raises one other fundamentally important issue about risk assessment. "The role of science is to present information about hazards," he says. "But it's the role of the decision-makers to take that information, and a lot of other information, in order to make decisions about public welfare." In fact the legal fight in L'Aquila is viewed by some as a philosophical dispute between scientists, who believe that their role is pure hazard assessment, and the local prosecutors, who argue that Italian law obliges scientific advisers to evaluate the fragility of buildings and other factors in their assessment of risk.

Scientists will also have to work hard to convince the public, at least in L'Aquila, that frequent, probabilistic risk assessment is a better way to protect them than age-old traditions. As Vittorini told Picuti after the earthquake, the messages from the commission meeting "may have in some way deprived us of the fear of earthquakes. The science, on this occasion, was dramatically superficial, and it betrayed the culture of prudence and good sense that our parents taught us on the basis of experience and of the wisdom of the previous generations."

Glancing at an image of his deceased wife and daughter on his mobile phone, Vittorini says: "We're not crazy people. We just want accountability. We hope this trial can be a symbol of change." ■ **SEE WORLD VIEW P. 251**

Stephen S. Hall is a science writer based in New York who also teaches public communication to graduate students in science at New York University.

➔ NATURE.COM
To hear Stephen S. Hall discussing the events in L'Aquila: go.nature.com/otcael

COMMENT

HISTORY Copernicus biography from Dava Sobel mixes fact and fiction **p.276**

MUSIC In conversation with climate-change composer Paul D. Miller **p.279**

CULTURE Martin Kemp muses on 15 years of artists in lab schemes **p.278**

CORRESPONDENCE Pros and cons of 24/7 working stir up debate **p.280**



D. ACKER/BLOOMBERG VIA GETTY



A drilling operation in Bradford County, Pennsylvania: one of the many places where shale rocks are fractured to release oil and gas.

Should fracking stop?

Extracting gas from shale increases the availability of this resource, but the health and environmental risks may be too high.

POINT

Yes, it's too high risk

Natural gas extracted from shale comes at too great a cost to the environment, say Robert W. Howarth and Anthony Ingraffea.

Natural gas from shale is widely promoted as clean compared with oil and coal, a 'win-win' fuel that can lessen emissions while still supplying abundant fossil energy over coming decades until a switch to renewable energy sources is made. But shale gas isn't clean, and shouldn't be used as a bridge fuel.

Shale rock formations can contain vast amounts of natural gas (which is mostly methane). Until quite recently, most of **PAGE 272 ►**

COUNTERPOINT

No, it's too valuable

Fracking is crucial to global economic stability; the economic benefits outweigh the environmental risks, says Terry Engelder.

After a career in geological research on one of the world's largest gas supplies, I am a born-again 'cornucopian'. I believe that there is enough domestic gas to meet our needs for the foreseeable future thanks to technological advances in hydraulic fracturing. According to IHS, a business-information company in Douglas County, Colorado, the estimated recoverable gas from US shale source rocks using fracking is about 42 trillion cubic metres, almost **PAGE 274 ►**

POINT: FRACKING: TOO HIGH RISK ▶ this gas was not economically obtainable, because shale is far less permeable than the rock formations exploited for conventional gas. Over the past decade or so, two new technologies have combined to allow extraction of shale gas: 'high-volume, slick-water hydraulic fracturing' (also known as 'fracking'), in which high-pressure water with additives is used to increase fissures in the rock; and precision drilling of wells that can follow the contour of a shale layer closely for 3 kilometres or more at depths of more than 2 kilometres (see 'Fracking for fuel'). Industry first experimented with these two technologies in Texas about 15 years ago. Significant shale-gas production in other states, including Arkansas, Pennsylvania and Louisiana, began only in 2007–09. Outside North America, only a handful of shale-gas wells have been drilled.

Industry sources claim that they have used fracking to produce more than 1 million oil and natural gas wells since the late 1940s. However, less than 2% of the well fractures since the 1940s have used the high-volume technology necessary to get gas from shale, almost all of these in the past ten years. This approach is far bigger and riskier than the conventional fracking of earlier years. An average of 20 million litres of water are forced under pressure into each well, combined with large volumes of sand or other materials to help keep the fissures open, and 200,000 litres of acids, biocides, scale inhibitors, friction reducers and surfactants. The fracking of a conventional well uses at

most 1–2% of the volume of water used to extract shale gas¹.

Many of the fracking additives are toxic, carcinogenic or mutagenic. Many are kept secret. In the United States, such secrecy has been abetted by the 2005 'Halliburton loophole' (named after an energy company headquartered in Houston, Texas), which exempts fracking from many of the nation's major federal environmental-protection laws, including the Safe Drinking Water Act. In a 2-hectare site, up to 16 wells can be drilled, cumulatively servicing an area of up to 1.5 square kilometres, and using 300 million litres or more of water and additives. Around one-fifth of the fracking fluid flows back up the well to the surface in the first two weeks, with more continuing to flow out over the lifetime of the well. Fracking also extracts natural salts, heavy metals, hydrocarbons and radioactive materials from the shale, posing risks to ecosystems and public health when these return to the surface. This flowback is collected in open pits or large tanks until treated, recycled or disposed of.

Because shale-gas development is so new, scientific information on the environmental costs is scarce. Only this year have studies begun to appear in peer-reviewed journals, and these give reason for pause. We call for a moratorium on shale-gas development to allow for better study of the cumulative risks to water quality, air quality and global climate. Only with such comprehensive knowledge can appropriate regulatory frameworks be developed.

We have analysed the well-to-consumer lifecycle greenhouse-gas

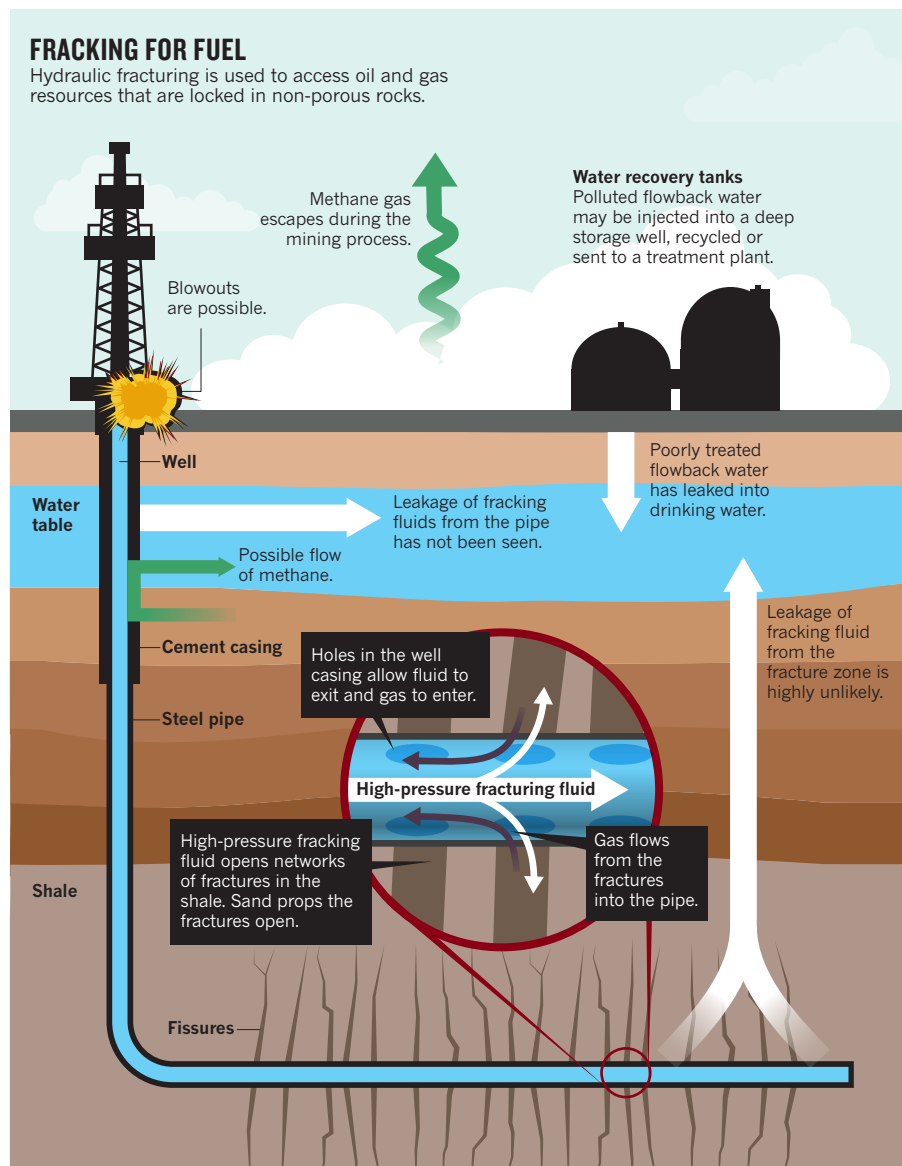
footprint of shale gas when used for heat generation (its main use), compared with conventional gas and other fossil fuels — the first estimate in the peer-reviewed literature². Methane is a major component of this footprint, and we estimate that 3.6–7.9% of the lifetime production of a shale gas well (compared with 1.7–6% for conventional gas wells) is vented or leaked to the atmosphere from the well head, pipelines and storage facilities. In addition, carbon dioxide is released both directly through the burning of the gas for heat, and to a lesser extent indirectly through the process of developing the resource.

Methane is a potent greenhouse gas, so even small emissions matter. Over a 20-year time period, the greenhouse-gas footprint of shale gas is worse than that for coal or oil (see 'A daunting climate footprint'). The influence of methane is lessened over longer time scales, because methane does not stay in the atmosphere as long as carbon dioxide. Still, over 100 years, the footprint of shale gas remains comparable to that of oil or coal.

When used to produce electricity rather than heat, the greater efficiency of gas plants compared with coal plants slightly lessens the footprint of shale gas³. Even then, the total greenhouse-gas footprint from shale gas exceed those of coal at timescales of less than about 50 years.

Methane venting and leakage can be decreased by upgrading old pipelines and storage systems, and by applying better technology for capturing gas in the 2-week flowback period after fracking. But current economic incentives are not sufficient to drive such improvements; stringent regulation will be required. In July, the US Environmental Protection Agency released a draft rule that would push industry to reduce at least some methane emissions, in part focusing on post-fracking flowback. Nonetheless, our analysis² indicates that the greenhouse-gas footprint of shale gas is likely to remain large.

Another peer-reviewed study looked at



private water wells near fracking sites⁴. It found that about 75% of wells sampled within 1 kilometre of gas drilling in the Marcellus shale in Pennsylvania were contaminated with methane from the deep shale formations. Isotopic fingerprinting of the methane indicated that deep shale was the source of contamination, rather than biologically derived methane, which was present at much lower concentrations in water wells at greater distances from gas wells. The study found no fracking fluids in any of the drinking-water wells examined. This is good news, because these fluids contain hazardous materials, and methane itself is not toxic. However, methane poses a high risk of explosion at the levels found, and it suggests a potential for other gaseous substances in the shale to migrate with the methane and contaminate water wells over time.

Have fracking-return fluids contaminated drinking water? Yes, although the evidence is not as strong as for methane contamination, and none of the data has yet appeared in the peer-reviewed literature (although a series of articles in *The New York Times* documents the problem, for example go.nature.com/58hxot and go.nature.com/58koj3). Contamination can happen through blowouts, surface spills from storage facilities, or improper disposal of fracking fluids. In Texas, flowback fluids are disposed of through deep injection into abandoned gas or oil wells. But such wells are not available everywhere. In New York and Pennsylvania, some of the waste is treated in municipal sewage plants that weren't designed to handle these toxic and radioactive wastes. Subsequently, there has been contamination of tributaries of the Ohio River with barium, strontium and bromides from municipal wastewater treatment plants receiving fracking wastes⁵. This contamination apparently led to the formation of dangerous brominated hydrocarbons in municipal drinking-water supplies that relied on these surface waters, owing to interaction of the contaminants with organic matter during the chlorination process.

Shale-gas development — which uses huge diesel pumps to inject the water — also creates local air pollution, often at dangerous levels. Volatile hydrocarbons such as benzene (which occurs naturally in shale, and is a commonly used fracking additive) are one major concern. The state of Texas reports benzene concentrations in air in the Barnett shale area that sometimes exceed acute toxicity standards⁶, and although the concentrations observed in the Marcellus shale area in Pennsylvania are lower⁷ (with only 2,349 wells drilled at the time these air contaminants were reported, out of an expected total of 100,000), they are high enough to pose a risk of cancer from chronic exposure⁸. Emissions from drills, compressors, trucks and other machinery can lead to very high levels of ground-level ozone, as documented in parts of Colorado that had not experienced severe air pollution before shale-gas development⁹.

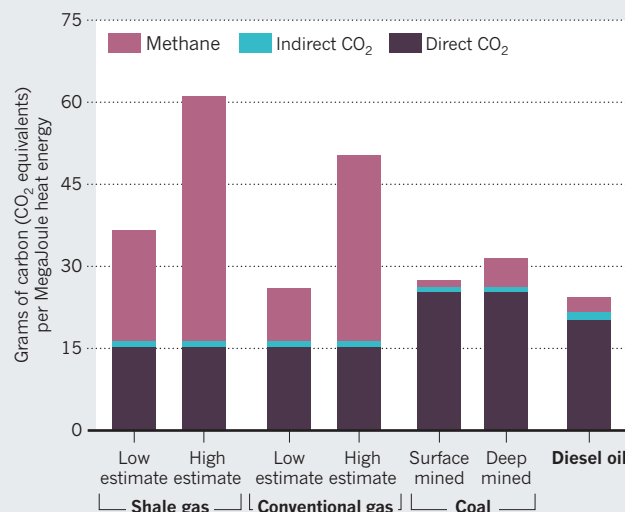
UNPROFITABLE PROGRESS

The argument for continuing shale-gas exploitation often hinges on the presumed gigantic size of the resource. But this may be exaggerated. The Energy Information Administration of the US Department of Energy estimates that 45% of US gas supply will come from shale gas by 2035 (with the vast majority of this replacing conventional gas, which has a lower greenhouse-gas footprint). Other gas industry observers are even more bullish. However, David Hughes, a geoscientist with more than 30 years experience with the Canadian Geological Survey, concludes in his report for the Post Carbon Institute, a non-profit group headquartered in Santa Rosa, California, that forecasts are likely to be overstated, perhaps greatly so³. Last month, the US Geological Survey released a new estimate of the amount of gas in the Marcellus shale formation (the largest shale-gas formation in the United States), concluding that the Department of Energy has overestimated the resource by some five-fold¹⁰.

Shale gas may not be profitable at current prices, in part because

A DAUNTING CLIMATE FOOTPRINT

Over 20 years, shale gas is likely to have a greater greenhouse effect than conventional gas or other fossil fuels.



production rates for shale-gas wells decline far more quickly than for conventional wells. Although very large resources undoubtedly exist in shale reservoirs, an unprecedented rate of well drilling and fracking would be required to meet the Department of Energy's projections, which might not be economic³. If so, the recent enthusiasm over shale gas could soon collapse, like the dot-com bubble.

Meanwhile, shale gas competes for investment with green energy technologies, slowing their development and distracting politicians and the public from developing a long-term sustainable energy policy.

With time, perhaps engineers can develop more appropriate ways to handle fracking-fluid return wastes, and perhaps the technology can be made more sustainable and less polluting in other ways. Meanwhile, the gas should remain safely in the shale, while society uses energy more efficiently and develops renewable energy sources more aggressively. ■

Robert W. Howarth is in the Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York, New York 14853, USA. **Anthony Ingraffea** is in the School of Civil and Environmental Engineering at Cornell University, Ithaca, New York, New York 14853, USA.
e-mail: rwh2@cornell.edu

1. New York State Department of Environmental Conservation *Revised Draft SGEIS on the Oil, Gas and Solution Mining Regulatory Program* (Sept. 2011); available at: <http://go.nature.com/yzponk>
2. Howarth, R. W., Santoro, R. & Ingraffea, A. *Clim. Change* **106**, 679–690 (2011).
3. Hughes, D. *Will Natural Gas Fuel America in the 21st Century?* (Post Carbon Institute, 2011); available at: <http://go.nature.com/gkboqm>
4. Osborn, S. G., Vengosh, A., Warner, N. R. & Jackson, R. B. *Proc. Natl Acad. Sci. USA* **108**, 8172–8176 (2011).
5. Volz, C. D. et al. *Contaminant Characterization of Effluent from Pennsylvania Brine Treatment Inc., Josephine Facility Being Released into Blacklick Creek, Indiana County, Pennsylvania* (2011); available at: <http://go.nature.com/5otd59>
6. Texas Commission on Environmental Quality. *Barnett Shale Formation Area Monitoring Projects* (2010); available at: <http://go.nature.com/v7k4re>
7. Pennsylvania Department of Environmental Protection. *Northeastern Pennsylvania Marcellus Shale Short-Term Ambient Air Sampling Report* (2011); available at: <http://go.nature.com/tjcsnt>
8. Talbott, E. O. et al. *Environ. Res.* **111**, 597–602 (2011).
9. Colorado Department of Public Health and Environment. *Public Health Implications of Ambient Air Exposures as Measured in Rural and Urban Oil & Gas Development Areas — an Analysis of 2008 Air Sampling Data* (2010); available at: <http://go.nature.com/5tttna>
10. Coleman, J. L. et al. *Assessment of Undiscovered Oil and Gas Resources of the Devonian Marcellus Shale of the Appalachian Basin Province, 2011*. US Geological Survey Fact Sheet 2011–3092 (2011); available at <http://go.nature.com/8kejhm>

COUNTERPOINT: FRACKING: TOO VALUABLE ► equal to the total conventional gas discovered in the United States over the past 150 years, and equivalent to about 65 times the current US annual consumption. During the past three years, about 50 billion barrels of additional recoverable oil have been found in shale oil deposits — more than 20% of the total conventional recoverable US oil resource. These ‘tight’ oil resources, which also require fracking to access, could generate 3 million barrels a day by 2020, offsetting one-third of current oil imports. International data aren’t as well known, but the effect of fracking on global energy production will be huge (see ‘Global gas reserves’).

Global warming is a serious issue that fracking-related gas production can help to alleviate. In a world in which productivity is closely linked to energy expenditure, fracking will be vital to global economic stability until renewable or nuclear energy carry more of the workload. But these technologies face persistent problems of intermittency and lack of power density or waste disposal. Mankind’s inexorable march towards 9 billion people will require a broad portfolio of energy resources, which can be gained only with breakthroughs such as fracking. Such breakthroughs should be promoted by policy that benefits the economy yet reduces overall greenhouse-gas emissions. Replacing coal with natural gas in power plants, for example, reduces the plants’ greenhouse emissions by up to 50% (ref. 1).

At present, fracking accounts for 50% of locally produced natural gas (see ‘US natural-gas production set to explode’) and 33% of local petroleum. The gas industry in America accounts for US\$385 billion in direct economic activity and nearly 3 million jobs. Because gas wells have notoriously steep production declines, stable supplies depend on a steady rate of new well completions. A moratorium on new wells would have an immediate and harsh effect on the US economy that would trigger a global ripple.

Global warming aside, there is no compelling environmental reason to ban hydraulic fracturing. There are environmental risks, but these

can be managed through existing, and rapidly improving, technologies and regulations. It might be nice to have moratoria after each breakthrough to study the consequences (including the disposal of old batteries or radioactive waste), but because energy expenditure and economic health are so closely linked, global moratoria are not practical.

The gains in employment, economics and national security, combined with the potential to reduce global greenhouse-gas emissions if natural gas is managed properly, make a compelling case.

NO NEED FOR PANIC

I grew up with the sights, sounds and smells of the Bradford oil fields in New York state. My parents’ small farm was over a small oil pool, with fumes from unplugged wells in the air and small oil seeps coating still waters. Before college, I worked these oil fields as a roustabout, mainly cleaning pipes and casings. Like me, most people living in such areas are not opposed to drilling, it seems. In my experience, such as during the recent hearings for the Pennsylvania Governor’s Marcellus Shale Advisory Commission, activists from non-drilling regions outnumber those from drilling regions by approximately two to one.

Modern, massive hydraulic fracturing is very different from that used decades ago. Larger pads are required to accommodate larger drill rigs, pumps and water supplies. People usually infer from this that modern techniques have a greater impact on the environment. This isn’t necessarily true. Although more water is used per well, there are far fewer wells per unit area. In the Bradford oil fields in the 1950s, a 640-acre parcel of land might have held more than 100 wells, requiring some 18 kilometres of roads, and with a lattice of surface pipelines. During the Marcellus development today, that same parcel of land is served by a single pad of five acres, with a 0.8-kilometre right-of-way for roads and pipelines.

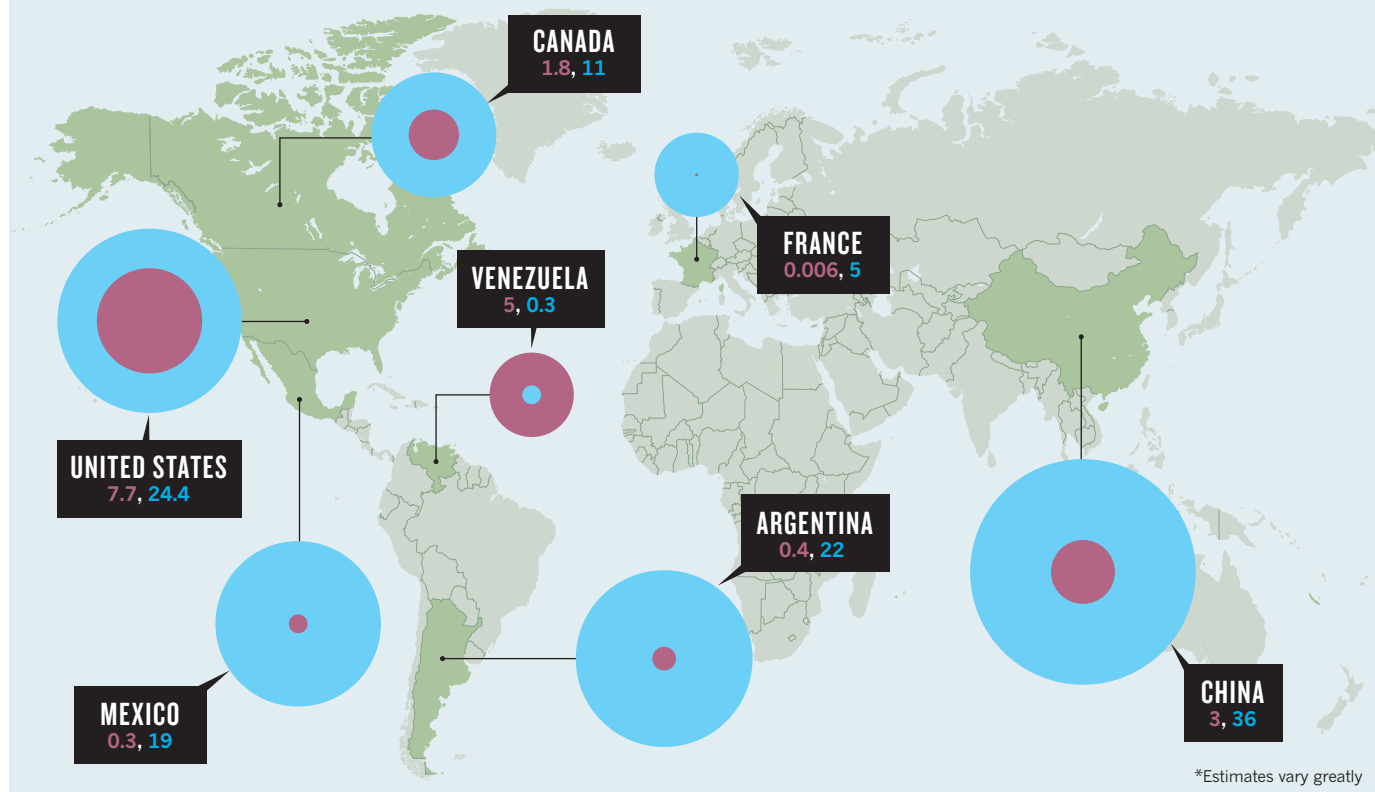
Although ‘fracking’ has emerged as a scare term in the press,

GLOBAL GAS RESERVES

Using fracking to access shale gas would vastly increase gas resources in many countries. Russia and the Middle East are not included because their large reserves of easily accessible gas will render shale gas less important there.

Proven gas reserves
(trillion cubic metres)

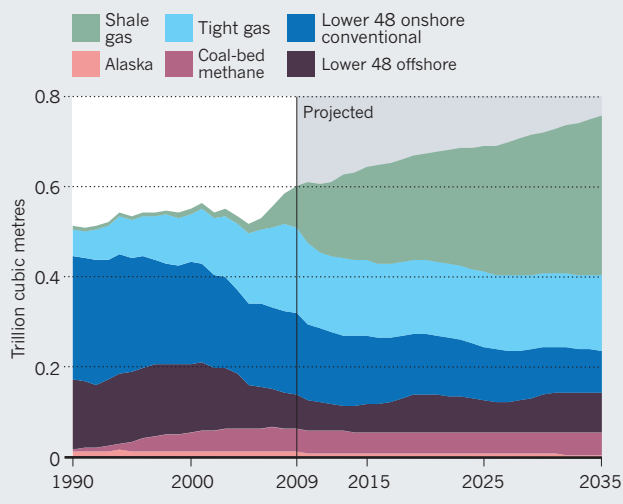
Technically recoverable shale gas resources*
(trillion cubic metres)



SOURCE: EIA

US NATURAL-GAS PRODUCTION SET TO EXPLODE

Shale-gas output already matches production from offshore wells in the lower 48 states (mainland US states excluding Alaska). Gas (shale and tight) extracted by fracking is set to overtake all other sources.



hydraulic fracturing is not so strange or frightening. The process happens naturally: high-pressure magma, water, petroleum and gases deep inside Earth can crack rock, helping to drive plate tectonics, rock metamorphism and the recycling of carbon dioxide between the mantle and the atmosphere.

Oil and gas have their origins in muds rich with organic matter in low-oxygen water bodies. Over millions of years, some of these deposits were buried and 'cooked' in the deep Earth, turning the organic matter into fossil fuel and the mud to shale rocks. In many areas, natural hydraulic fracturing allowed a large portion of oil and gas to escape from the dense, impermeable shale and migrate into neighbouring, more porous rocks. Some of this fossil fuel was trapped by cap rock, creating the conventional reserves that mankind has long tapped. The groundwater above areas that host such conventional deposits naturally contains methane, thanks to natural hydraulic fracturing of the rock and the upward seeping of gas into the water table over long time periods.

More than 96% of all oil and gas has been released from its original source rocks; industrial hydraulic fracturing aims to mimic nature to access the rest. As in nature, industrial fracking can be done with a wide variety of gases and liquids. Nitrogen can be used to open cracks in the shale, for example. But this is inefficient, because of the energy lost by natural decompression of the nitrogen gas. Water is more efficient, because very little energy is wasted in decompression. Sand is added to prop open the cracks, and compounds such as surface-tension reducers are added to improve gas recovery.

UNDER CONTROL

Two main environmental concerns are water use and water contamination. Millions of gallons of water are required to stimulate a well. In Pennsylvania, high rainfall means that water is abundant, and regulations ensure that operators stockpile rainwater during the wet season to use during drier months (thus the injection of massive volumes of water in the Bradford oil fields for secondary recovery of oil, once the well pressure has fallen, flew under the radar of environmentalists for half a century). Obtaining adequate water for industrial fracking in dry regions such as the Middle East and western China is a local concern, but is no reason for a global moratorium.

Press reports often repeat strident concerns about the chemicals added to fracking fluids. But many of these compounds are relatively benign. One commonly used additive is similar to simethicone, which is also used in antacids to reduce surface tension and turn small bubbles in the stomach into larger ones that can move along more easily.

Many of the industrial additives are common in household products. Material safety data sheets for these additives are required by US regulation. Industry discloses additives on a website called FracFocus.org, run by state regulators.

Some people have expressed worries that fracking fluids might migrate more than 2 kilometres upwards from the cracked shale into groundwater. The Ground Water Protection Council, a non-profit national association of state groundwater and underground-injection control agencies headquartered in Oklahoma City, has found no instance in which injected fluid contaminated groundwater from below². This makes sense: water cannot flow this distance uphill in timescales that matter. This is the premise by which deep disposal wells, used to hold toxic waste worldwide, are considered safe. During gas production, the pressure of methane is reduced: this promotes downward, not upward flow of these fluids.

Gas shale contains a number of materials that are carried back up the pipe to the surface in flowback water, including salts of barium and radioactive isotopes, that might be harmful in concentrated form. According to a recent *New York Times* analysis, these elements can be above the US Environmental Protection Agency's sanctioned background concentrations in some flowback tanks. Industry is moving towards complete recycling of these fluids so this should be of less concern to the public. However, production water will continue to flow to the surface in modest volumes throughout the life of a well; this water needs to be, and currently is, treated to ensure safe disposal.

The real risk of water contamination comes from these flowback fluids leaking into streams or seeping down into groundwater after reaching the surface. This can be caused by leaky wellheads, holding tanks or blowouts. Wellheads are made sufficiently safe to prevent this eventuality; holding tanks can be made secure; and blowouts, while problematic, are like all accidents caused by human error — an unpredictable risk with which society lives.

"With hydraulic fracturing, as in many cases, fear levels exceed the evidence."

Although methane coming up to the surface within the steel well pipe cannot escape into the surrounding rocks or groundwater, it is possible that the cement seal between the well and the bedrock might allow methane from shallow sandstone layers (rather than the reservoir deep below) to seep up into groundwater. Methane is a tasteless and odourless component of groundwater that can be consumed without ill effect when dissolved. It is not a poison. Long before gas-shale drilling, regulators warned that enclosed spaces, such as houses, should be properly ventilated in areas with naturally occurring methane in groundwater.

An alarm has been sounded too about the effect of escaped methane on global warming. The good news is that methane has a very short half-life in the atmosphere: carbon dioxide emitted during the building of the first Sumerian cities is still affecting our climate, whereas escaped methane from the fracturing of the Barnett shale in 1997 is more than half gone. Industry can and should take steps to reduce air emissions, by capturing or flaring methane and converting motors and compressors from diesel to natural gas.

Risk perception is ultimately subjective: facts are all too easily combined with emotional responses. With hydraulic fracturing, as in many cases, fear levels exceed the evidence. ■

Terry Engelder is in the department of geosciences at Pennsylvania State University, University Park, Pennsylvania 16802, USA.
e-mail: jte2@psu.edu

1. Jiang, M. *Environ. Res. Lett.* **6**, 034014 (2011).
2. Statement of Scot Kell, on behalf of the Ground Water Protection Council, to the House Committee on Natural Resources Subcommittee on Energy and Mineral Resources (4 June 2009); available at: <http://go.nature.com/5jl2bp>

The author declares competing financial interests: details accompany this article online at go.nature.com/pjenyww.



The spark that ignited Nicolaus Copernicus' interest in the positions of heavenly spheres remains a mystery.

ASTRONOMY

Recasting the heavens

Dava Sobel mixes fact and fiction to great effect in her biography of Copernicus, finds **Owen Gingerich**.

Nicolaus Copernicus (1473–1543) was a quintessential unifier. He gathered the planets around a central Sun, thereby inventing the Solar System. The grand challenge facing biographers of Copernicus is that few personal details survive. Was he gregarious? Did he have a girlfriend in his youth? When did he become interested in the stars? Because the historical record answers none of these questions, some authors have resorted to fiction. In *A More Perfect Heaven*, science historian Dava Sobel has it both ways.

The first third of the book is strictly historical. It traces Copernicus's life from his birth in Polish Toruń, to his undergraduate studies in Kraków and graduate work in Italy, and to his career as a legal, medical and administrative officer in the northernmost Catholic diocese of Poland. Sobel makes particularly effective use of the records surviving from when Canon Copernicus was the rent collector for the Frauenburg cathedral chapter's vast land holdings. To fill the narrative, she occasionally quotes from John

"The grand challenge facing biographers of Copernicus is that few personal details survive."

Banville's novel *Doctor Copernicus* (W. W. Norton, 1976), but always makes clear that those reconstructions are not taken from archival sources. What Sobel

achieves is a brilliant chronological account — the best in the literature — laying out the stages of Copernicus's administrative career.

Having reached Copernicus's final years, when publication of his still unfinished manuscript seemed to have stalled, Sobel takes a new tack. She inserts a play in 2 acts and 17 scenes, with 6 characters. Besides Copernicus, there is the ailing and rabidly anti-Lutheran bishop, Johannes Dantiscus. And there is Tiedemann Giese, a more liberal bishop of the adjacent diocese, who is Copernicus's confidant. The play opens with a young mathematics teacher from Lutheran Wittenberg, Georg Joachim Rheticus, literally stumbling in to seek an audience with Copernicus. True to the historical record, Rheticus finally persuades the ageing canon to allow a copy of his manuscript to be taken to Nuremberg for printing. The penultimate scene brings the printed pages of *De Revolutionibus Orbium Coelestium* (*On the Revolutions of the Heavenly Spheres*) to the astronomer's deathbed. In contrast to this reasonably well documented event, the final fictional scene shows Rheticus turning up at Copernicus's funeral, and receiving as a gift the original manuscript.

The invented scenes paint vividly the peril that faced a young Lutheran

NATURE.COM

A museum of Copernicus's instruments: go.nature.com/gz4jnx

J.-L. HUEENS/NATIONAL GEOGRAPHIC STOCK



A More Perfect Heaven: How Copernicus Revolutionized the Cosmos

DAVA SOBEL
Walker/Bloomsbury:
2011. 288 pp.
\$25/£14.99

clandestinely at work with Copernicus in Catholic territory. The introduction of Anna, Copernicus's housekeeper — whom Bishop Dantiscus believed to be a harlot — as a fifth character provides another point of tension partially documented in the surviving correspondence. The sixth and entirely fictitious character, the young acolyte Franz, serves as a focus for Rheticus's only partly concealed homosexuality. Sobel's literary handling of these issues gives a dramatic punch to an otherwise colourless encounter between Copernicus and Rheticus, his only student — an encounter that was crucial to placing the heliocentric cosmology on the world stage.

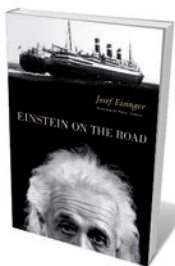
The final third of the book returns to documented history, tracing the aftermath of the 1543 publication of *De Revolutionibus*, which led to the insights of Johannes Kepler and Galileo Galilei, and the eventual acceptance of the Copernican system. The wonderful detail and eloquent writing that Sobel demonstrated in her best-selling *Longitude* and *Galileo's Daughter* carry the reader along here too. Given what she has chosen to include, the book is first rate.

What *A More Perfect Heaven* does not include are questions that have puzzled historians of science for many decades. What triggered Copernicus's interest in the radical heliocentric arrangement? "With a wave of his hand, he had made the Earth a planet and set it spinning," she writes elegantly. There is, however, barely a clue as to what *De Revolutionibus* contains, or how one might use it to calculate positions of the Sun and the planets. Nor is there much about Copernicus as an observer. His manuscript contained only 30 or so new observations, but they are crucial points for establishing the parameters of the planets' orbits, and Copernicus had to wait for years before some of the desired astronomical configurations took place.

A More Perfect Heaven is a charming and accurate book, although it omits much of the technical background in which earlier accounts revelled. Still, this carefully constructed biography leaves space for those of us probing the origins of heliocentrism to defend our speculations. ■

Owen Gingerich is professor emeritus of astronomy and history of science at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, USA.
e-mail: ginger@cfa.harvard.edu

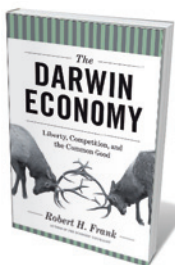
Books in brief



Einstein on the Road

Josef Eisinger PROMETHEUS 270 pp. \$25 (2011)

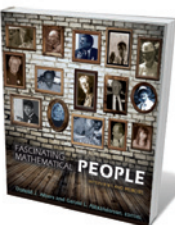
The 1920s and early 1930s saw Albert Einstein gripped by wanderlust. Prompted by Germany's political upheaval and a curiosity about other cultures — as well as a liking for contemplation on the high seas — the scientific celebrity roved from Japan to Uruguay and from California to Britain, encountering such luminaries as Charlie Chaplin, Niels Bohr, Edwin Hubble and Franklin Delano Roosevelt on the way. Einstein's travelogues form the core of physicist Josef Eisinger's portrait, an account that brings to life the artistic and scientific revolutions that were then in full swing.



The Darwin Economy: Liberty, Competition, and the Common Good

Robert H. Frank PRINCETON UNIVERSITY PRESS 256 pp. \$26.95 (2011)

The premise of economist Adam Smith's 'invisible hand' — a tenet of market economics — is that competitive self-interest shunts benefits to the community. But that is the exception rather than the rule, argues writer Robert H. Frank. Charles Darwin's idea of natural selection is a more accurate reflection of how economic competition works, he says, because individual and species benefits do not always coincide. Highlighting reasons for market failure and the need to cut waste, Frank argues that we can domesticate our wild economy by taxing higher-end spending and harmful industrial emissions.



Fascinating Mathematical People: Interviews and Memoirs

Edited by Donald J. Albers and Gerald L. Alexanderson PRINCETON UNIVERSITY PRESS 352 pp. \$35 (2011)

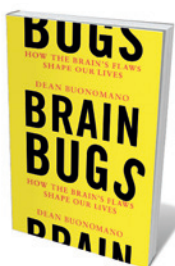
What do a Beatles expert, a professional magician and a Los Angeles dentist have in common? If they're Joseph Gallian, Arthur Benjamin and Leon Bankoff, it's mathematics. The words of these and other researchers, mentors and teachers in the maths community feature in this compilation by educator Donald Albers and mathematician Gerald Alexanderson. There is much to relish in these accounts — not least geometer Thomas Banchoff's friendship with Salvador Dalí, who explored the nexus of atomic science, maths and art late in life.



The Art of Medicine: Over 2,000 Years of Images and Imagination

Julie Anderson, Emm Barnes and Emma Shackleton UNIVERSITY OF CHICAGO PRESS 256 pp. \$50 (2011)

Portrayals of our grapplings with disease pop up throughout history. Medical historian Julie Anderson, with science communicators Emm Barnes and Emma Shackleton, survey a range of works from London's Wellcome Collection that highlight medical practices, including paintings, anatomical drawings, scrolls and digital art. Two millennia of visual exploration from cultures such as ancient Persia and Renaissance Europe provide a stunning overview of how ideas about healing the body and mind have evolved.



Brain Bugs: How the Brain's Flaws Shape Our Lives

Dean Buonomano W. W. NORTON 240 pp. £16.99 (2011)

Neurobiologist Dean Buonomano reframes the brain as a glitch-ridden lump of neural 'wetware' that often gets in the way of well-being. Information saturation in the man-made environment may threaten to overwhelm our brains' capabilities. But by getting to grips with its 'bugs' — including a vulnerability to advertising, gambling, fears, phobias and beliefs, an unreliable memory and a predilection for immediate gratification — we can uncover solutions to strengthen key mental functions, he says.



A More Perfect Heaven: How Copernicus Revolutionized the Cosmos

DAVA SOBEL
Walker/Bloomsbury:
2011. 288 pp.
\$25/£14.99

clandestinely at work with Copernicus in Catholic territory. The introduction of Anna, Copernicus's housekeeper — whom Bishop Dantiscus believed to be a harlot — as a fifth character provides another point of tension partially documented in the surviving correspondence. The sixth and entirely fictitious character, the young acolyte Franz, serves as a focus for Rheticus's only partly concealed homosexuality. Sobel's literary handling of these issues gives a dramatic punch to an otherwise colourless encounter between Copernicus and Rheticus, his only student — an encounter that was crucial to placing the heliocentric cosmology on the world stage.

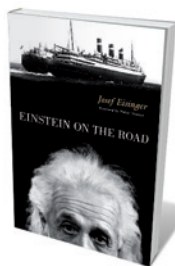
The final third of the book returns to documented history, tracing the aftermath of the 1543 publication of *De Revolutionibus*, which led to the insights of Johannes Kepler and Galileo Galilei, and the eventual acceptance of the Copernican system. The wonderful detail and eloquent writing that Sobel demonstrated in her best-selling *Longitude* and *Galileo's Daughter* carry the reader along here too. Given what she has chosen to include, the book is first rate.

What *A More Perfect Heaven* does not include are questions that have puzzled historians of science for many decades. What triggered Copernicus's interest in the radical heliocentric arrangement? "With a wave of his hand, he had made the Earth a planet and set it spinning," she writes elegantly. There is, however, barely a clue as to what *De Revolutionibus* contains, or how one might use it to calculate positions of the Sun and the planets. Nor is there much about Copernicus as an observer. His manuscript contained only 30 or so new observations, but they are crucial points for establishing the parameters of the planets' orbits, and Copernicus had to wait for years before some of the desired astronomical configurations took place.

A More Perfect Heaven is a charming and accurate book, although it omits much of the technical background in which earlier accounts revelled. Still, this carefully constructed biography leaves space for those of us probing the origins of heliocentrism to defend our speculations. ■

Owen Gingerich is professor emeritus of astronomy and history of science at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, USA.
e-mail: ginger@cfa.harvard.edu

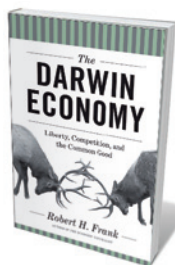
Books in brief



Einstein on the Road

Josef Eisinger PROMETHEUS 270 pp. \$25 (2011)

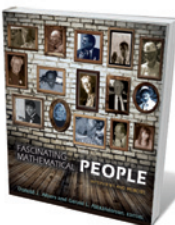
The 1920s and early 1930s saw Albert Einstein gripped by wanderlust. Prompted by Germany's political upheaval and a curiosity about other cultures — as well as a liking for contemplation on the high seas — the scientific celebrity roved from Japan to Uruguay and from California to Britain, encountering such luminaries as Charlie Chaplin, Niels Bohr, Edwin Hubble and Franklin Delano Roosevelt on the way. Einstein's travelogues form the core of physicist Josef Eisinger's portrait, an account that brings to life the artistic and scientific revolutions that were then in full swing.



The Darwin Economy: Liberty, Competition, and the Common Good

Robert H. Frank PRINCETON UNIVERSITY PRESS 256 pp. \$26.95 (2011)

The premise of economist Adam Smith's 'invisible hand' — a tenet of market economics — is that competitive self-interest shunts benefits to the community. But that is the exception rather than the rule, argues writer Robert H. Frank. Charles Darwin's idea of natural selection is a more accurate reflection of how economic competition works, he says, because individual and species benefits do not always coincide. Highlighting reasons for market failure and the need to cut waste, Frank argues that we can domesticate our wild economy by taxing higher-end spending and harmful industrial emissions.



Fascinating Mathematical People: Interviews and Memoirs

Edited by Donald J. Albers and Gerald L. Alexanderson PRINCETON UNIVERSITY PRESS 352 pp. \$35 (2011)

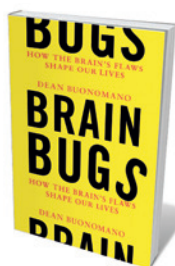
What do a Beatles expert, a professional magician and a Los Angeles dentist have in common? If they're Joseph Gallian, Arthur Benjamin and Leon Bankoff, it's mathematics. The words of these and other researchers, mentors and teachers in the maths community feature in this compilation by educator Donald Albers and mathematician Gerald Alexanderson. There is much to relish in these accounts — not least geometer Thomas Banchoff's friendship with Salvador Dalí, who explored the nexus of atomic science, maths and art late in life.



The Art of Medicine: Over 2,000 Years of Images and Imagination

Julie Anderson, Emm Barnes and Emma Shackleton UNIVERSITY OF CHICAGO PRESS 256 pp. \$50 (2011)

Portrayals of our grapplings with disease pop up throughout history. Medical historian Julie Anderson, with science communicators Emm Barnes and Emma Shackleton, survey a range of works from London's Wellcome Collection that highlight medical practices, including paintings, anatomical drawings, scrolls and digital art. Two millennia of visual exploration from cultures such as ancient Persia and Renaissance Europe provide a stunning overview of how ideas about healing the body and mind have evolved.



Brain Bugs: How the Brain's Flaws Shape Our Lives

Dean Buonomano W. W. NORTON 240 pp. £16.99 (2011)

Neurobiologist Dean Buonomano reframes the brain as a glitch-ridden lump of neural 'wetware' that often gets in the way of well-being. Information saturation in the man-made environment may threaten to overwhelm our brains' capabilities. But by getting to grips with its 'bugs' — including a vulnerability to advertising, gambling, fears, phobias and beliefs, an unreliable memory and a predilection for immediate gratification — we can uncover solutions to strengthen key mental functions, he says.

Artists in the lab

Martin Kemp explores the nature of science–art collaborations after 15 years of major initiatives around the world.

Caged canaries were placed in coal mines to warn of poisonous gases in the twentieth century. Programmes to insert artists into laboratories proliferated in the 1990s, just as the canaries had been phased out. Do these ‘artists-in-residence’ act as metaphorical canaries, detecting practices that are potentially noxious? Or are they cuddly creators, obedient poodles who translate scientists’ work into publicly accessible forms?

Two of the largest science–art schemes raise most of the big questions and provide some answers about the nature of these collaborations. The first, the Sciart grant scheme of the Wellcome Trust, Britain’s leading biomedical charity, was introduced in 1996 and has lent its name to the whole area of activity. The second is a laboratory, equipped with experimental apparatus and organic materials but staffed by artists. The SymbioticA research centre at the University of Western Australia, Perth, was inaugurated by artists Oron Catts and Ionat Zurr as the Tissue Culture & Art Project (TC&A) in 1996, and was formally incorporated into the university as SymbioticA in 2000. It marked its tenth anniversary with a provocative exhibition at the Science Gallery of Trinity College Dublin (see *Nature* 470, 334; 2011).

BRIDGING THE GAP

The intellectual breeze that fanned the flame of initiatives such as these arose from a concern among many artists and scientists that the divorce between their disciplines was unhealthy. Without subscribing to the moribund aspects of C. P. Snow’s 1959 diagnosis of the ‘Two Cultures’, it was easy to agree that mid-twentieth-century art and science had become dangerously isolated from each other and from society at large (see *Nature* 459, 32; 2009). Stories in the press reinforced the perceived weirdness of artists and scientists in the public mind. Notorious images such as the mouse bearing a cartilaginous human-like ear on its back, created by the scientist Charles Vacanti, and the green fluorescent rabbit engineered by the artist Eduardo Kac, fed Frankenstein fantasies in the public imagination.

The Wellcome scheme proved that there was a substantial demand for bridge-building. After a decade, it had received nearly 1,500 applications and made 124 awards totalling

almost £3 million (US\$4.8 million). Some projects were initiated by scientists, others by artists. In 2006, the scheme was replaced by an extended programme with an investment of £1.2 million a year.

Funding in the sciences and arts is usually formalized around predetermined programmes, standard research protocols and predictable outcomes. Risk has been leached out. The Wellcome Trust rightly recognized that unproven collaborations between imaginative scientists and creative artists needed a different approach at every stage. Risk had to be embraced. The trick was to identify a special creative chemistry between high-level participants.

Although Ken Arnold, the administrator of the scheme, admitted after two years that it wasn’t obvious what the ingredients were for success, he and subsequent evaluators now agree that most of the funded projects have had positive outcomes. Many have resulted in exhibited art of high quality or have generated scientific, social, cultural, economic and personal gains for participants and the public (see the Wellcome Trust’s report at go.nature.com/xz6gnb).

Defining the gains for scientists has proved more elusive than evaluating how artists have benefited. In a gratifying number of instances, host scientists reported that they had acquired broader perspectives on their work or its communication. The artistic presentation of their research in galleries and public spaces has proved salutary for them. Once an image is in the public domain, strict management of its reception is no longer possible, and that can be discomforting and educational. Good artists are expert in this slippery domain and have much to teach the scientists.

ASYMMETRIES

Asymmetries abound in these collaborations. The projects matter in professional terms far more to the artists than the scientists. Little, if any, kudos is to be gained by the scientist in having a Sciart project on his or her CV. It would be good if scientists received more recognition for their participation. For the artist, the collaboration can be an important career move, opening up new venues and audiences.

Participating scientists tend to be well enough established not to have to worry about ‘wasting time’ on an art project. They are often older, male and of high status. Large numbers of the artists are female, young and aspiring.

Some projects are marred by a scientist’s belief that he or she can enjoy becoming an artist. It is usually taken as read that the artists will not become professional research scientists during this brief spell. It is a commentary on a general view of artists that the reverse is not seen to hold true. Art, like science, requires highly specialized skills honed over long periods of education and experience.

The grants involved — mostly in the region of £30,000, with a few greatly exceeding

this — are substantial for the arts but relatively minor for a successful lab. Much of the money in the early days of the Sciart scheme went towards costs, with artists receiving little or no payment for a great deal of hard work. By contrast, the scientists are likely to be in receipt of a regular salary. Recently, the Wellcome Trust has endeavoured to



Semi-Living Worry Doll A covered by living cells (left) grown on a 2-cm-high scaffold (right).

THE TISSUE CULTURE & ART PROJECT, SEMI-LIVING WORRY DOLL A (2000)/ARS ELECTRONICA

ensure that the artists receive adequate remuneration.

The SymbioticA model is different. As residents in their own lab, the artists there have the same academic status as experimental scientists on campus. The lab competes for funding within the university and outside. The experimental apparatus and materials are used in a scientific manner, but the resulting research is not published in the way that a scientist would recognize. SymbioticA's greatest achievements have been to establish a different institutional model and attitude towards the end products.

ARTISTIC CONCERNS

The project *Semi-Living Worry Dolls*, by Catts and Zurr, still working under the name of TC&A, is as much about the process and its recording as it is about fixed artistic products. Traditional worry dolls are given to Guatemalan children so that they can share their concerns with a trusted confidant. The dolls by TC&A are confected from degradable polymers and surgical sutures. The polymers are progressively replaced by living cells within a micro-gravity bioreactor.

First exhibited in Linz, Austria, in 2000, the dolls were the first tissue-engineered sculptures to be presented alive in a gallery. Viewers are invited to speak their worries to the dolls into an adjacent microphone. The anonymous responses have gone further than the anticipated concerns about biological engineering; visitors often spoke about personal issues. SymbioticA's style of artwork is about process and participation, not an enduring material object.

Art-science collaboration is becoming established as a distinct curatorial practice that has a defined public engagement through exhibitions. Educational initiatives are arising, ranging from school programmes to master of arts degrees, such as the two-year postgraduate course at the University of the Arts in London. The notion of artists and scientists collaborating is no longer a surprise, and is a well recognized strategy in the art world.

As the Wellcome and SymbioticA examples show, artists in laboratories come to understand the science in such a way that they act as neither canaries nor poodles in a crudely critical or acquiescent manner. At their best, the artists present works of complexity and subtlety that engage the spectator's imagination in a non-prescriptive way. Ultimately, as with all artworks, the artist lays down the melody while encouraging the visitors to sing their songs in their own way. ■

Martin Kemp is emeritus professor of art history at the University of Oxford, UK.

Q&A Paul D. Miller Climate-change DJ

Paul D. Miller, also known as DJ Spooky, is famed for his digital sampling techniques. His 2007 foray to Antarctica inspired a multimedia symphony, Terra Nova: Sinfonia Antarctica, and a companion volume, The Book of Ice. Ahead of a performance of Terra Nova this week at the New York Academy of Sciences, he discusses how he uses weather patterns in his compositions.

How did you become an audio artist?

It was a hobby gone out of control. As a kid I messed around with early Texas Instruments and Commodore 64 computers. My mother made me take violin and double-bass lessons. After college, where I majored in philosophy and French literature, I started DJ'ing to pay my rent, which freed me up for writing and artwork. I began using digital sampling as a kind of musical collage, like the 'cut-up' text technique of Beat Generation author William S. Burroughs.

Why did you go to Antarctica in 2007?

I challenged myself to travel to one of the most remote parts of the planet and make acoustic portraits there. I wanted to confront the recursive logic of weather patterns — rain, snow, ice and wind. So I chartered a decommissioned Russian military ice-breaker ship and went to the continent.

How did you gather material for *Terra Nova*?

I carried a compact recording studio in a backpack across the ice. I set up microphones to record the sounds of water and ice, took photographs and distilled a composition from them, mixing electronic edits of the sounds with string arrangements. I wanted to turn weather patterns, which are so complex it takes a supercomputer to model them, into audio-visual compositions. My aim was to convey the idea that, with climate change, some natural variables are no longer meshing.

How did *The Book of Ice* come about?

The book started as a graphical score for the musical piece, inspired by the work of British experimental composer Cornelius Cardew. It grew into a larger project: to condense the complex information about Antarctica into a digestible format using graphic design. String theorist Brian Greene, of Columbia University in New York, wrote a foreword about the physics of ice. And the book includes an infographic on the interactions between different causes of climate change.

What intrigues you about Antarctica?

It is the only continent with no government. One could think of it as a creative commons. A 1959 treaty forbids a military presence. The United States and others have put a huge amount of money into science there, and



M. FIGGIS

The Art of Climate Science: Antarctica

New York Academy of Sciences, New York. 7 p.m., 19 September.

The Book of Ice

PAUL D. MILLER
Mark Batty: 2011.
128 pp. \$29.95

the research scene has a military feel. Fortunately, the scientists share information with colleagues from other countries.

You have also started an artists' centre on Vanuatu. Why?

The Pacific island of Vanuatu keeps getting ranked as one of the happiest places on Earth. My centre there pulls artists out of the city and slows them down. I've also worked on Nauru, a Pacific dystopia. After the Soviet Union collapsed, Nauru was an offshore banking centre, with billions of dollars passing through daily. It was economically devastated when the money vanished. I made recordings there and used them in a string-quartet composition and visual installation called *The Nauru Elegies*.

What's next?

My composition *Arctic Rhythms* is set at the North Pole. I travelled last year to the Svalbard archipelago. There are some 20 million people in the Arctic Circle and about 2,000 in Antarctica. A bigger population makes for a different project: it is about local frameworks, nation states, the international rule of law and the human response to climate change.

What's your view of climate change now?

Economists try to assign a cost to global warming. Yet biologist Richard Dawkins' theory of 'extended phenotype' says that anything an animal makes can be considered an effect of its genes on the environment. So we need to start thinking of climate change as an extension of what it means to be human. ■

INTERVIEW BY JASCHA HOFFMAN

ensure that the artists receive adequate remuneration.

The SymbioticA model is different. As residents in their own lab, the artists there have the same academic status as experimental scientists on campus. The lab competes for funding within the university and outside. The experimental apparatus and materials are used in a scientific manner, but the resulting research is not published in the way that a scientist would recognize. SymbioticA's greatest achievements have been to establish a different institutional model and attitude towards the end products.

ARTISTIC CONCERNS

The project *Semi-Living Worry Dolls*, by Catts and Zurr, still working under the name of TC&A, is as much about the process and its recording as it is about fixed artistic products. Traditional worry dolls are given to Guatemalan children so that they can share their concerns with a trusted confidant. The dolls by TC&A are confected from degradable polymers and surgical sutures. The polymers are progressively replaced by living cells within a micro-gravity bioreactor.

First exhibited in Linz, Austria, in 2000, the dolls were the first tissue-engineered sculptures to be presented alive in a gallery. Viewers are invited to speak their worries to the dolls into an adjacent microphone. The anonymous responses have gone further than the anticipated concerns about biological engineering; visitors often spoke about personal issues. SymbioticA's style of artwork is about process and participation, not an enduring material object.

Art-science collaboration is becoming established as a distinct curatorial practice that has a defined public engagement through exhibitions. Educational initiatives are arising, ranging from school programmes to master of arts degrees, such as the two-year postgraduate course at the University of the Arts in London. The notion of artists and scientists collaborating is no longer a surprise, and is a well recognized strategy in the art world.

As the Wellcome and SymbioticA examples show, artists in laboratories come to understand the science in such a way that they act as neither canaries nor poodles in a crudely critical or acquiescent manner. At their best, the artists present works of complexity and subtlety that engage the spectator's imagination in a non-prescriptive way. Ultimately, as with all artworks, the artist lays down the melody while encouraging the visitors to sing their songs in their own way. ■

Martin Kemp is emeritus professor of art history at the University of Oxford, UK.

Q&A Paul D. Miller Climate-change DJ

Paul D. Miller, also known as DJ Spooky, is famed for his digital sampling techniques. His 2007 foray to Antarctica inspired a multimedia symphony, Terra Nova: Sinfonia Antarctica, and a companion volume, The Book of Ice. Ahead of a performance of Terra Nova this week at the New York Academy of Sciences, he discusses how he uses weather patterns in his compositions.

How did you become an audio artist?

It was a hobby gone out of control. As a kid I messed around with early Texas Instruments and Commodore 64 computers. My mother made me take violin and double-bass lessons. After college, where I majored in philosophy and French literature, I started DJ'ing to pay my rent, which freed me up for writing and artwork. I began using digital sampling as a kind of musical collage, like the 'cut-up' text technique of Beat Generation author William S. Burroughs.

Why did you go to Antarctica in 2007?

I challenged myself to travel to one of the most remote parts of the planet and make acoustic portraits there. I wanted to confront the recursive logic of weather patterns — rain, snow, ice and wind. So I chartered a decommissioned Russian military ice-breaker ship and went to the continent.

How did you gather material for *Terra Nova*?

I carried a compact recording studio in a backpack across the ice. I set up microphones to record the sounds of water and ice, took photographs and distilled a composition from them, mixing electronic edits of the sounds with string arrangements. I wanted to turn weather patterns, which are so complex it takes a supercomputer to model them, into audio-visual compositions. My aim was to convey the idea that, with climate change, some natural variables are no longer meshing.

How did *The Book of Ice* come about?

The book started as a graphical score for the musical piece, inspired by the work of British experimental composer Cornelius Cardew. It grew into a larger project: to condense the complex information about Antarctica into a digestible format using graphic design. String theorist Brian Greene, of Columbia University in New York, wrote a foreword about the physics of ice. And the book includes an infographic on the interactions between different causes of climate change.

What intrigues you about Antarctica?

It is the only continent with no government. One could think of it as a creative commons. A 1959 treaty forbids a military presence. The United States and others have put a huge amount of money into science there, and



M. FIGGIS

The Art of Climate Science: Antarctica

New York Academy of Sciences, New York. 7 p.m., 19 September.

The Book of Ice

PAUL D. MILLER
Mark Batty: 2011.
128 pp. \$29.95

the research scene has a military feel. Fortunately, the scientists share information with colleagues from other countries.

You have also started an artists' centre on Vanuatu. Why?

The Pacific island of Vanuatu keeps getting ranked as one of the happiest places on Earth. My centre there pulls artists out of the city and slows them down. I've also worked on Nauru, a Pacific dystopia. After the Soviet Union collapsed, Nauru was an offshore banking centre, with billions of dollars passing through daily. It was economically devastated when the money vanished. I made recordings there and used them in a string-quartet composition and visual installation called *The Nauru Elegies*.

What's next?

My composition *Arctic Rhythms* is set at the North Pole. I travelled last year to the Svalbard archipelago. There are some 20 million people in the Arctic Circle and about 2,000 in Antarctica. A bigger population makes for a different project: it is about local frameworks, nation states, the international rule of law and the human response to climate change.

What's your view of climate change now?

Economists try to assign a cost to global warming. Yet biologist Richard Dawkins' theory of 'extended phenotype' says that anything an animal makes can be considered an effect of its genes on the environment. So we need to start thinking of climate change as an extension of what it means to be human. ■

INTERVIEW BY JASCHA HOFFMAN

Correspondence

Review boards: all need closer scrutiny

As director of a consulting group that works with institutional review boards (IRBs) at universities, hospitals and commercial organizations in the United States, I disagree that commercial IRBs are unduly influenced by profits and are less thorough than their academic counterparts (*Nature* 476, 125; 2011).

Many of the IRBs enveloped in your critique are accredited by the Association for the Accreditation of Human Research Protection Programs (AAHRPP). The two IRBs censured by the US Food and Drug Administration, Essex and Coast, were not. AAHRPP accreditation is voluntary; organizations undergo a rigorous assessment of their policies and records, including on-site interviews to ensure compliance with federal regulations and AAHRPP standards.

For accreditation, IRBs must separate business decisions from ethical review, even though this is not federally mandated. For instance, independent IRBs accredited by the AAHRPP prohibit equity holders from serving as IRB members or participating in research review. Independent IRBs are constantly evaluated by sponsors, clinical research organizations, regulators and the AAHRPP. Profit is often reinvested in training in ethics and regulatory processes for IRB members and staff.

Absent from your Editorial was an acknowledgement that universities and hospitals can have a proprietary interest in their research; independent IRBs do not. I believe that additional scrutiny of all IRBs is needed: protecting human subjects in research overseen by a hospital or university IRB is just as important as protecting those in research reviewed by independent IRBs.

Nicholas C. Slack HRP Consulting Group, Rockville, Maryland, USA. slackn@thehrpconsultinggroup.com

Competing financial interests declared. See go.nature.com/m5qaue.

South Korean energy plan is unrealistic

South Korea imports 97% of its energy and is the world's tenth-largest emitter of greenhouse gases. It has increased its target for supplying renewable energy from 2.4% in 2008 to 6.1% by 2020. This seems overly ambitious, given that its renewable energy has increased by only 0.37% in the past decade. Even that aim is modest compared with the European Union's goal to source 20% of its energy from renewables by 2020.

South Korea is attempting to transform from quantitative to low-carbon qualitative growth (*Nature* 464, 832–833; 2010). This green-growth strategy encourages policies that tackle climate change and enhance security, and aims to create new markets by investing 2% of gross domestic product in renewable-energy sources over the next five years.

These measures are unrealistic, however, given the state of the South Korean new- and renewable-energy industry. Even with an export boom, the country's lack of original technology and facilities could result in profits going overseas. The problem lies with South Korea's high dependence on imports of core components for export goods, combined with its sluggish rate of change to domestic production.

Scientists must agree on which new- and renewable-energy technologies are suitable for adoption. They need to take into account economic factors, convenience, safety and reliability, and to convince industry and consumers to recognize the advantages.

Hyung-Man Kim Inje University, Gimhae, South Gyeongsang, South Korea. mechkhm@inje.ac.kr

NATURE'S READERS COMMENT ONLINE

A taste of the lively discussion on working 24/7 (Nature 477, 5, 20–22 and 27–28; 2011).

Kausik Datta says:

Hard work is essential, but most major scientific discoveries are arrived at by serendipity, the appreciation of which requires creativity and a thinking, enlightened mind. A slave-driving mentorship that encourages drone-like devotion to work and assembly-line productivity will only result in early burn out and the loss of love for science.

kdatta1@jhmi.edu

Jessica Mark Welch says:

Science demands hard work, but to sacrifice your health and your family life, so that while nominally spending time with your kids you are on the phone with your lab? How unreconstructed. I do not want a world where only people who can live that way can be scientists.

jmarkwelch@mbl.edu

Burkhard Haefner says:

All of us need time to relax and think or even to dream — to let the soul dangle, as we say in German. We all know the story of Isaac Newton wasting away his time, or so it seemed, lying under an apple tree.

bhaefner@its.jnj.com

Dean Griffiths says:

Rarely do insights occur after 14 hours of picking colonies. While it may be great for a PI [principal investigator] to publish lots of mediocre papers, students and postdocs require big papers to become established — and constantly working insane hours is unlikely to achieve this. Plus there really are times with your family that you can never get back. Is it worth missing

them to do another PCR?

dsg29@cam.ac.uk

Maya Capelson says:

An average life scientist in the lab, grad student or postdoc, working 50–60 hours per week, will probably produce at least one paper in 4–5 years. Twenty-seven people working over 100 hours a week [in the lab profiled] produce just 29 papers in 5–6 years. So pretty much the same productivity as a scientist working for only half that time.

capelson@salk.edu

Chris Wood says:

I respect Alfredo Quiñones-Hinojosa for his honesty and the fact that he screens out his applicants to ensure they fully realize what they are getting into. And if they cannot stay the pace, he supports them in transferring somewhere more appropriate.

chris@ibt.unam.mx

Srikrishna Pandey says:

Some researchers and engineers really enjoy their work, so when they have to work overtime it doesn't occur to them to resent it.

srikrishnapandey@gmail.com

Julien Marquis says:

Some PIs may never have experienced the devastation of trashing a year's work. It is important and even pleasant to work very hard, but not always and not on anything. So PIs — if you want your crew (particularly naive PhD students) to work hard, ensure that they are pursuing a promising track.

julien.marquis@epfl.ch

To join this debate, go to go.nature.com/djydh.

EVOLUTION

Selection for positive illusions

Everybody knows that overconfidence can be foolhardy. But a study reveals that having an overly positive self-image might confer an evolutionary advantage if the rewards outweigh the risks. [SEE LETTER P.317](#)

MATTHIJS VAN VEELEN & MARTIN A. NOWAK

Ask anyone with a driver's licence to rate their own abilities behind the wheel, and most people will report that they are above average¹. The same is true for self-assessments of performance in cognitive tasks², of attractiveness³ (by men, not by women) and of the healthiness of our behaviour⁴: people typically place themselves higher on the ladder than they really are. In a survey of 1 million high-school students⁵, a solid 70% rated themselves as above-average leaders (versus 2% who thought of themselves as below average), and a spectacular 94% of college professors possess teaching abilities that are above average — according to themselves⁶.

Obviously they cannot all be right, but that does not make them dysfunctional or mentally unhealthy. In fact, one way to get self-assessments to obey some minimal aggregate consistency is to restrict surveys to sufficiently depressed people⁷ (although this finding has been questioned^{8,9}). Mentally healthy people blissfully suffer from what are called positive illusions: they overestimate their abilities, as well as their control over events, and they underestimate their vulnerability to risk¹⁰. Of course, one can overrate oneself too much, as do sufferers from narcissistic personality disorder or megalomania, but healthy people's estimates of their own abilities seem to start just a little above where they really are. Reporting on page 317 of this issue, Johnson and Fowler¹¹ describe a model that might explain why this is so.

An obvious question is how overconfidence survives the process of natural selection. The prevalence of rose-tinted self-assessments suggests that it might even be adaptive to be overconfident — in contrast to schizophrenia, for instance, which is maladaptive but nonetheless exists in moderate proportions in humans. But how can it be adaptive to misjudge how you compare with others? You would think that an incorrect assessment of one's own capabilities can induce only misguided decisions.

One suggested explanation is that there is a benefit in having others think that you're great. And as there is no better way of being a strong persuader than firmly believing in yourself, this would lead to an upward bias in how



GETTY IMAGES

Figure 1 | Float like a butterfly, sting like a bee. Muhammad Ali saw himself as “the king of the world”. His supreme confidence helped him to win many fights. Johnson and Fowler¹¹ report that overconfidence can confer an evolutionary advantage.

people perceive themselves compared with others¹². That may lead to a mistake here and there, but the benefits of the esteem of others could outweigh that (Fig. 1).

Johnson and Fowler¹¹ suggest a remarkable alternative explanation. According to their model, a biased self-belief can actually lead people to make the right decision, whereas an unbiased self-image would lead to a suboptimal decision. That sounds counter-intuitive, but the key lies in the authors' departure from what could be called the ‘naïve economist's’ idea of how humans arrive at decisions (‘naïve’ because many economists are not that naïve at all).

The authors' model envisages a valuable resource that two individuals can decide to claim or not. If both claim it, then they will fight over it — which is costly for both. The stronger individual will win the fight and gain access to the resource. If only one of them claims the resource, it goes to that person. If neither claims it, no one gets it.

Now if both contenders could simply assess the fighting strength of the other with perfect accuracy, the optimal strategy would be a

no-brainer: fight if you are stronger, concede if you are weaker. But it gets interesting if the contestants have imperfect information about each other's strength. In this situation, contestants might back off because they think their opponent is stronger than he or she really is. A weaker contestant could then win a reward if she claims it while the opponent backs off.

This situation can be dealt with within the realm of what economists call perfect rationality, which assumes that both parties understand all aspects of their situation, and that they correctly anticipate the odds that the other player will claim the resource. But Johnson and Fowler suggest that there is a short cut to the right decision. The short cut combines a simple heuristic — fight if you think you're stronger — with a bias. If the resource is valuable relative to the cost of fighting, then the risk of an extra battle here and there is outweighed by the gains made when otherwise unclaimed resources are won, which makes overestimating one's own fighting abilities worthwhile. If the cost of fighting is large relative to the value of the resource, then it is better to underestimate one's own strength. The behaviours

described by the authors' model are actually more complex than described above, because the model also predicts that populations can, for instance, evolve to a stable mixture of both over- and under-confident people.

Another evolutionary explanation is the following: overconfidence could reduce average pay-off, but top performers will still come from the group of overconfident individuals. For example, overconfidence about roulette-playing 'abilities' will lead to overall losses from this game, but the best performers will have played often. Strong selection — as in 'winner takes all' — should favour overconfidence.

Johnson and Fowler's study¹¹ prompts a variety of interesting questions. The 'winning strategy' (for low fighting costs) can be wired into the brain in two ways. The first involves a simple heuristic plus overconfidence: only fight when you think you are stronger, but overestimate your strength. The second way involves perfect rationality without overconfidence: given some uncertainty, the winning strategy can be to fight opponents even if they seem slightly stronger than you. Future empirical and theoretical studies might help to decide

which of these two describes us best.

It would also be interesting to establish a link between the authors' findings and overconfidence in trading behaviour¹³, the willingness to buy overly complex financial products (which are thought to have led to the current crisis in the banking system), political decisions that lead to war¹⁴, and the evolution of fighting behaviour in animals¹⁵. Given that 94% of college professors rate themselves as above average, there should be enough overconfidence around to tackle all the natural follow-up questions. ■

Matthijs van Veelen is at the Center for Research in Experimental Economics and Political Decision Making, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands. **Martin A. Nowak** is at the Program for Evolutionary Dynamics, Department of Mathematics and Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. e-mails: c.m.vanveelen@uva.nl; martin_nowak@harvard.edu

1. Svenson, O. *Acta Psychol.* **47**, 143–148 (1981).
2. Kruger, J. & Dunning, D. *J. Pers. Social Psychol.* **77**, 1121–1134 (1999).
3. Gabriel, M. T., Critelli, J. W. & Ee, J. S. *J. Pers.* **62**, 143–155 (1994).
4. Hoorens, V. & Harris, P. *Psychol. Health* **13**, 451–466 (1998).
5. Alicke, M. D. & Govorun, O. in *The Self in Social Judgment* (eds Alicke, M. D., Dunning, D. A. & Krueger, J. I.) 85–106 (Psychology, 2005).
6. Cross, P. *New Directions Higher Educ.* **17**, 1–15 (1977).
7. Taylor, S. E. & Brown, J. D. *Psychol. Bull.* **103**, 193–210 (1988).
8. Shedler, J. et al. *Am. Psychol.* **48**, 1117–1131 (1993).
9. Colvin, C. R. & Block, J. *Psychol. Bull.* **116**, 3–20 (1994).
10. Sharot, T. *The Optimism Bias: A Tour of the Irrationally Positive Brain* (Pantheon, 2011).
11. Johnson, D. D. P. & Fowler, J. H. *Nature* **477**, 317–320 (2011).
12. Trivers, R. *Deceit and Self-Deception: Fooling Yourself the Better to Fool Others* (Allen Lane, 2011).
13. Barber, B. M. & Odean, T. *Q. J. Econ.* **116**, 261–292 (2001).
14. Johnson, D. D. P. *Overconfidence and War: The Havoc and Glory of Positive Illusions* (Harvard Univ. Press, 2004).
15. Enquist, M. & Leimar, O. *J. Theor. Biol.* **127**, 187–205 (1987).

CELL DIVISION

Six degrees of separation

During cell division, the DNA-associated CENP-A protein recruits the kinetochore protein complex to assemble on chromosomes. A region of just six amino-acid residues earmarks CENP-A for this purpose. SEE LETTER P.354

ALISON PIDOUX & ROBIN ALLSHIRE

In chromosomes, a series of protein particles act like spools, packaging DNA into a structure called chromatin. The spools are known as nucleosomes, and most are composed of eight subunits — two subunits of each of the four major histone proteins, H2A, H2B, H3 and H4. However, the centromeric sites, which are required for chromosome segregation during cell division, are different. Instead of two subunits of H3, centromeric nucleosomes contain the centromere-specific histone H3 variant called CENP-A. Two papers^{1,2}, including one by Guse *et al.* on page 354 of this issue, provide structural and mechanistic insights into the workings of CENP-A-containing nucleosomes.

At cell division, chromosome segregation is orchestrated by the kinetochore — a complex machinery composed of more than 100 proteins through which chromosomes attach to the microtubules that form the spindle apparatus, which allows chromosome segregation. In most

eukaryotes (organisms such as animals, plants and fungi), kinetochores are assembled at centromeres. Centromeres frequently contain extensive arrays of repetitive DNA sequences, such as the 100–10,000-kilobase repeats in the 'α-satellite' DNA family found in human centromeres. Kinetochores assemble on only a subset of these repeats, indicating that factors other than primary DNA sequence influence the site of assembly^{3,4}.

It is well known that heritable changes in genome function can occur through alterations that are independent of the DNA sequence — a process referred to as epigenetic propagation. Epigenetic phenomena are frequently mediated by post-translational modification of histones through the addition of chemical entities such as acetyl and methyl groups, which form epigenetic 'marks'. Such marks promote the assembly of specific chromatin states that are crucial for many cellular and developmental processes. CENP-A itself has an extreme epigenetic character, and, by replacing histone H3, it provides a pivotal

mark for the formation of centromeres at a particular location on chromosomes^{3,4}.

Previous work⁵ in fruitfly cells showed that overexpression of CENP-A leads to the assembly of kinetochores at new sites, suggesting that CENP-A nucleosomes act alone to form a platform for kinetochore formation. Nonetheless, similar experiments on cultured human cells⁶ did not induce abnormal localization of kinetochores. To investigate how CENP-A directly effects the interaction between the centromere and kinetochores, Guse *et al.*¹ generated *in vitro* arrays of CENP-A nucleosomes assembled on DNA.

The authors find that, when placed in frog egg extracts, CENP-A nucleosomes can recruit kinetochore proteins. However, it remains unclear whether these structures contain the full repertoire of components associated with native centromeres, or whether they can mediate processes such as chromosome movement along microtubules. Nevertheless, Guse and colleagues' synthetic kinetochores clearly show aspects of normal kinetochore function: they display enhanced microtubule binding, and they seem to sense interactions with microtubules, eliciting a response that is indicative of an operational spindle-assembly checkpoint — the surveillance mechanism that ensures accurate chromosome segregation during cell division. Thus, in this *in vitro* system at least, CENP-A nucleosomes are sufficient to dictate 'functional' kinetochore assembly, whereas H3 nucleosomes assembled on the same DNA sequence are not. In other words, incorporating CENP-A in place of H3 makes the crucial difference that allows kinetochore formation *in vitro*.

described by the authors' model are actually more complex than described above, because the model also predicts that populations can, for instance, evolve to a stable mixture of both over- and under-confident people.

Another evolutionary explanation is the following: overconfidence could reduce average pay-off, but top performers will still come from the group of overconfident individuals. For example, overconfidence about roulette-playing 'abilities' will lead to overall losses from this game, but the best performers will have played often. Strong selection — as in 'winner takes all' — should favour overconfidence.

Johnson and Fowler's study¹¹ prompts a variety of interesting questions. The 'winning strategy' (for low fighting costs) can be wired into the brain in two ways. The first involves a simple heuristic plus overconfidence: only fight when you think you are stronger, but overestimate your strength. The second way involves perfect rationality without overconfidence: given some uncertainty, the winning strategy can be to fight opponents even if they seem slightly stronger than you. Future empirical and theoretical studies might help to decide

which of these two describes us best.

It would also be interesting to establish a link between the authors' findings and overconfidence in trading behaviour¹³, the willingness to buy overly complex financial products (which are thought to have led to the current crisis in the banking system), political decisions that lead to war¹⁴, and the evolution of fighting behaviour in animals¹⁵. Given that 94% of college professors rate themselves as above average, there should be enough overconfidence around to tackle all the natural follow-up questions. ■

Matthijs van Veelen is at the Center for Research in Experimental Economics and Political Decision Making, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands. **Martin A. Nowak** is at the Program for Evolutionary Dynamics, Department of Mathematics and Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. e-mails: c.m.vanveelen@uva.nl; martin_nowak@harvard.edu

1. Svenson, O. *Acta Psychol.* **47**, 143–148 (1981).
2. Kruger, J. & Dunning, D. *J. Pers. Social Psychol.* **77**, 1121–1134 (1999).
3. Gabriel, M. T., Critelli, J. W. & Ee, J. S. *J. Pers.* **62**, 143–155 (1994).
4. Hoorens, V. & Harris, P. *Psychol. Health* **13**, 451–466 (1998).
5. Alicke, M. D. & Govorun, O. in *The Self in Social Judgment* (eds Alicke, M. D., Dunning, D. A. & Krueger, J. I.) 85–106 (Psychology, 2005).
6. Cross, P. *New Directions Higher Educ.* **17**, 1–15 (1977).
7. Taylor, S. E. & Brown, J. D. *Psychol. Bull.* **103**, 193–210 (1988).
8. Shedler, J. et al. *Am. Psychol.* **48**, 1117–1131 (1993).
9. Colvin, C. R. & Block, J. *Psychol. Bull.* **116**, 3–20 (1994).
10. Sharot, T. *The Optimism Bias: A Tour of the Irrationally Positive Brain* (Pantheon, 2011).
11. Johnson, D. D. P. & Fowler, J. H. *Nature* **477**, 317–320 (2011).
12. Trivers, R. *Deceit and Self-Deception: Fooling Yourself the Better to Fool Others* (Allen Lane, 2011).
13. Barber, B. M. & Odean, T. *Q. J. Econ.* **116**, 261–292 (2001).
14. Johnson, D. D. P. *Overconfidence and War: The Havoc and Glory of Positive Illusions* (Harvard Univ. Press, 2004).
15. Enquist, M. & Leimar, O. *J. Theor. Biol.* **127**, 187–205 (1987).

CELL DIVISION

Six degrees of separation

During cell division, the DNA-associated CENP-A protein recruits the kinetochore protein complex to assemble on chromosomes. A region of just six amino-acid residues earmarks CENP-A for this purpose. SEE LETTER P.354

ALISON PIDOUX & ROBIN ALLSHIRE

In chromosomes, a series of protein particles act like spools, packaging DNA into a structure called chromatin. The spools are known as nucleosomes, and most are composed of eight subunits — two subunits of each of the four major histone proteins, H2A, H2B, H3 and H4. However, the centromeric sites, which are required for chromosome segregation during cell division, are different. Instead of two subunits of H3, centromeric nucleosomes contain the centromere-specific histone H3 variant called CENP-A. Two papers^{1,2}, including one by Guse *et al.* on page 354 of this issue, provide structural and mechanistic insights into the workings of CENP-A-containing nucleosomes.

At cell division, chromosome segregation is orchestrated by the kinetochore — a complex machinery composed of more than 100 proteins through which chromosomes attach to the microtubules that form the spindle apparatus, which allows chromosome segregation. In most

eukaryotes (organisms such as animals, plants and fungi), kinetochores are assembled at centromeres. Centromeres frequently contain extensive arrays of repetitive DNA sequences, such as the 100–10,000-kilobase repeats in the 'α-satellite' DNA family found in human centromeres. Kinetochores assemble on only a subset of these repeats, indicating that factors other than primary DNA sequence influence the site of assembly^{3,4}.

It is well known that heritable changes in genome function can occur through alterations that are independent of the DNA sequence — a process referred to as epigenetic propagation. Epigenetic phenomena are frequently mediated by post-translational modification of histones through the addition of chemical entities such as acetyl and methyl groups, which form epigenetic 'marks'. Such marks promote the assembly of specific chromatin states that are crucial for many cellular and developmental processes. CENP-A itself has an extreme epigenetic character, and, by replacing histone H3, it provides a pivotal

mark for the formation of centromeres at a particular location on chromosomes^{3,4}.

Previous work⁵ in fruitfly cells showed that overexpression of CENP-A leads to the assembly of kinetochores at new sites, suggesting that CENP-A nucleosomes act alone to form a platform for kinetochore formation. Nonetheless, similar experiments on cultured human cells⁶ did not induce abnormal localization of kinetochores. To investigate how CENP-A directly effects the interaction between the centromere and kinetochores, Guse *et al.*¹ generated *in vitro* arrays of CENP-A nucleosomes assembled on DNA.

The authors find that, when placed in frog egg extracts, CENP-A nucleosomes can recruit kinetochore proteins. However, it remains unclear whether these structures contain the full repertoire of components associated with native centromeres, or whether they can mediate processes such as chromosome movement along microtubules. Nevertheless, Guse and colleagues' synthetic kinetochores clearly show aspects of normal kinetochore function: they display enhanced microtubule binding, and they seem to sense interactions with microtubules, eliciting a response that is indicative of an operational spindle-assembly checkpoint — the surveillance mechanism that ensures accurate chromosome segregation during cell division. Thus, in this *in vitro* system at least, CENP-A nucleosomes are sufficient to dictate 'functional' kinetochore assembly, whereas H3 nucleosomes assembled on the same DNA sequence are not. In other words, incorporating CENP-A in place of H3 makes the crucial difference that allows kinetochore formation *in vitro*.

NATURE

50 Years Ago

A Collection of Tables and Nomograms for the Processing of Observations made on Artificial Earth Satellites. By I. D. Zhongolovich and V. M. Amlin — This volume is a welcome addition to the very few existing mathematical tables which are designed to help in the calculation of artificial satellite orbits ... The tables themselves have been reproduced from the Russian original by a photographic process, and only the 10 introductory pages needed translating. Despite its minimal contribution to the subject-matter, the Pergamon Press has chosen to charge £5 for the volume, although it was advertised at 70s ... There are irritating errors ... Also, the translation is at times sadly deficient ... These shortcomings do not, however, seriously mar the value of the volume to the specialist.

From *Nature* 16 September 1961

100 Years Ago

On Saturday, September 9, the aërial post was inaugurated by Mr. Gustav Hamel, one of our most brilliant flyers, who carried a sack of letters in a Blériot monoplane from Hendon to Windsor in thirteen minutes ... The other aviators who should have started were prevented by the thirty-mile wind, and no further deliveries took place until Monday, when Messrs. Greswell and Driver carried six mail-bags over in the early morning ... The affair has aroused great interest, so much so that it is as well to sound a word of warning and say that the aëroplane post is neither practical, useful, nor economical. Letters can be sent far more cheaply, trustworthily, and conveniently by train or motor-van, and it is to be expected that these conditions will continue for the next half-century at least ... Besides, it is unthinkable for very many years to come that we should put good aviators to the menial task of carrying mails regularly.

From *Nature* 14 September 1911

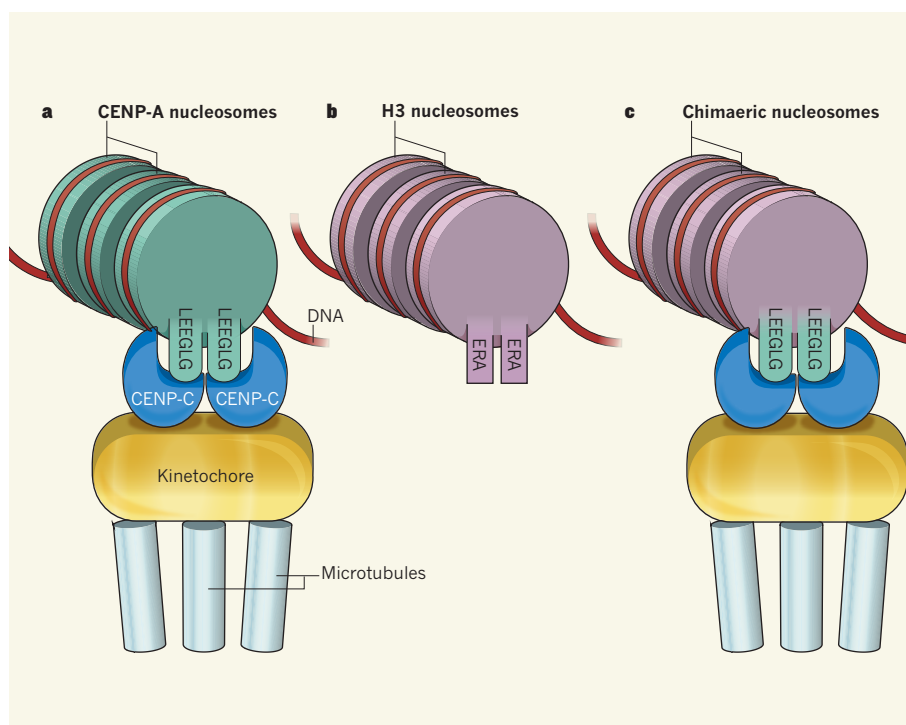


Figure 1 | Recognition of CENP-A-containing nucleosomes. **a**, Guse *et al.*¹ show that arrays of CENP-A nucleosomes allow active kinetochore-like structures that bind microtubules in frog egg extracts to assemble *in vitro*. This interaction is mediated by the centromeric protein CENP-C and requires the -LEEGLG amino-acid sequence at the carboxy terminus of CENP-A. **b**, H3 nucleosomes, which lack this particular motif and instead have three other residues (-ERA), do not assemble synthetic kinetochores. **c**, Replacement of -ERA at the carboxy terminus of H3 with -LEEGLG allows binding of CENP-C and assembly of synthetic kinetochores.

How do CENP-A nucleosomes differ from their H3 counterparts? Previous biophysical analyses hinted that the composition of CENP-A and H3 nucleosomes was radically different⁷. They suggested that CENP-A nucleosomes contain only a single subunit of H2A, H2B, CENP-A and H4, forming 'hemisomes' with half the height of typical nucleosomes.

In July, Tashiwana *et al.*² presented a high-resolution crystal structure of the human CENP-A nucleosome assembled *in vitro* on its natural substrate, α -satellite DNA. The structure shows that, like their H3 counterparts, *in vitro*-assembled CENP-A nucleosomes consist of eight subunits — with a CENP-A–H4 tetramer forming the core and bounded by two H2A–H2B dimers. But a clear difference is that CENP-A has a shorter helical domain at its amino terminus than H3. This alters DNA interactions at the nucleosome entry and exit points, meaning that CENP-A nucleosomes associate with a shorter stretch of DNA (121 compared with 147 base pairs). Moreover, two other amino-acid residues in one region (loop 1) cause it to jut out more in CENP-A nucleosomes, and their deletion affects CENP-A retention at, but not its targeting to, centromeres.

What attracts kinetochore proteins to CENP-A nucleosomes but not to H3 nucleosomes? Most differences between the properties

of CENP-A and H3 *in vivo* have been ascribed to a structural domain of CENP-A called CATD. This domain of CENP-A shows 22 differences in amino-acid residues compared with the same domain in H3 across a region of 40 residues. So, when placed within H3, it allows the resulting CENP-A–H3 chimaeric protein to be incorporated at centromeres⁴.

In Guse and colleagues' *in vitro* system¹, the CATD domain of CENP-A–H3 chimaeric proteins in nucleosome arrays was not sufficient or necessary for kinetochore assembly. Remarkably, when the authors replaced only the three carboxy-terminal residues of H3 (-ERA) with the six from CENP-A (-LEEGLG), this sequence allowed H3 to assemble kinetochore structures *in vitro* (Fig. 1). Further investigation showed that another centromeric protein, CENP-C, binds the -LEEGLG motif of human CENP-A — but only in the context of a nucleosome — to provide the platform for kinetochore formation. Presumably, once established, recognition of CENP-A nucleosomes, and their retention and replenishment at centromeres through subsequent cell divisions, are dictated by interactions with other kinetochore-associated proteins that may recognize additional differences such as loop 1 and CATD.

The *in vitro* formation of 'active' kinetochore complexes on purely CENP-A-containing nucleosomes¹ indicates that no neighbouring

H3 nucleosomes are required. H3 nucleosomes may, however, promote further stabilizing interactions between other kinetochore proteins and microtubules *in vivo*⁸. The conclusion that a key role of CENP-A chromatin is to provide a platform for kinetochore assembly is underscored by recent observations^{6,9,10} that the requirement for CENP-A for functional kinetochore formation can be bypassed by tethering the CENP-A chaperone protein HJURP — or other kinetochore proteins — to arrays of DNA binding sites.

The synthetic kinetochore-formation system described here¹ therefore provides

a new setting in which to further dissect the interactions required for the assembly of kinetochores on CENP-A chromatin. It will also allow investigation of the changes that occur in the composition of centromeric chromatin and kinetochores in response to cell-cycle events. ■

Alison Pidoux and Robin Allshire are at the Wellcome Trust Centre for Cell Biology, Institute of Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JR, UK.
e-mail: robin.allshire@ed.ac.uk

1. Guse, A., Carroll, C. W., Moree, B., Fuller, C. J. & Straight, A. F. *Nature* **477**, 354–358 (2011).
2. Tachiwana, H. *et al. Nature* **476**, 232–235 (2011).
3. Allshire, R. C. & Karpen, G. H. *Nature Rev. Genet.* **9**, 923–937 (2008).
4. Black, B. E. & Cleveland, D. W. *Cell* **144**, 471–479 (2011).
5. Olszak, A. M. *et al. Nature Cell Biol.* **13**, 799–808 (2011).
6. Gascoigne, K. E. *et al. Cell* **145**, 410–422 (2011).
7. Dalal, Y., Wang, H., Lindsay, S. & Henikoff, S. *PLoS Biol.* **5**, e218 (2007).
8. Screpanti, E. *et al. Curr. Biol.* **21**, 391–398 (2011).
9. Kiermaier, E. *et al. Nature Cell Biol.* **11**, 1109–1115 (2009).
10. Barnhart, M. C. *et al. J. Cell Biol.* **194**, 229–243 (2011).

EARTH SCIENCE

Lethal volcanism

Data from the Siberian Traps volcanic region suggest that its magma source includes a significant component of recycled oceanic crust. This finding helps to explain why basalt eruptions are so environmentally devastating. [SEE LETTER P.312](#)

PAUL B. WIGNALL

Large igneous provinces (LIPs) are the most spectacular manifestation of volcanism on Earth. They consist of huge individual basaltic lava flows, with volumes measured in thousands of cubic kilometres, stacked layer upon layer to form vast volcanic plateaux. No one has ever witnessed a LIP eruption, which is probably just as well because geologists suspect that they cause mass extinctions and major perturbations to the carbon cycle. However, until now our understanding of the link between LIPs and environmental change has been poor. A new hypothesis from the Sobolev brothers, Stephan and Alexander, and their colleagues¹ (page 312 of this issue) may provide the answer.

Gas emissions are the obvious culprit when it comes to looking for a volcanism-triggered global catastrophe. Sulphur dioxide (SO₂) and halogens both cause short-term cooling and acid rain, with carbon dioxide — another abundant volcanic gas — being the well-known greenhouse gas. Clearly, there is scope for volcanic gases to cause environmental damage, and we can start to evaluate this potential by calculating the volumes emitted during a LIP eruption. This is straightforward: first measure how much gas is released during a modern basalt-lava eruption (in Hawaii, for example) and then simply scale up the figures to the size of a LIP province, which is typically several million cubic kilometres.

This calculation produces some very large numbers, but these are values for total gas emissions spread over the lifetime of the province, which may be a million years². This point is important because the gases added to

the atmosphere are scrubbed out within a few years, leaving ample time for recovery between eruptions. Even CO₂ is significantly drawn down within a few thousand years, consumed during rock weathering, especially weathering of the basalt flows themselves³.

Because of this rapid atmospheric recovery time, environmental damage must be done either by individual eruptions or by a series

of closely spaced ones. This assertion is supported by recent studies⁴ revealing that the onset of LIP eruptions coincides precisely with extinction events — the first punch does all the damage. But there is a problem. The volumes of gas emitted per individual lava flow are not especially impressive; they are less than the yearly anthropogenic pollution fluxes of gases such as SO₂ and CO₂ (ref. 2). So, here is the rub: LIP eruptions coincide with environmental crises and yet simple calculation of gas fluxes suggests that their impact should be minor.

Sobolev and colleagues' hypothesis¹ concerns the source of LIP magma and proposes that it is much more gas-rich than is generally assumed, especially at the onset of eruptions. Traditional dogma has LIPs originating from deep-generated mantle plumes that tap primitive material which has not undergone previous



Figure 1 | Siberian flood-basalt flows in Putorana, Taymyr Peninsula. These rocks form part of a gigantic volcanic area that erupted 250 million years ago contemporaneously with the end-Permian mass extinction. A new study by Sobolev *et al.*¹ suggests that incorporation of large amounts of ocean-crust material in the erupting magma would have generated huge volumes of gases capable of triggering this environmental calamity.

NATURE

50 Years Ago

A Collection of Tables and Nomograms for the Processing of Observations made on Artificial Earth Satellites. By I. D. Zhongolovich and V. M. Amlin — This volume is a welcome addition to the very few existing mathematical tables which are designed to help in the calculation of artificial satellite orbits ... The tables themselves have been reproduced from the Russian original by a photographic process, and only the 10 introductory pages needed translating. Despite its minimal contribution to the subject-matter, the Pergamon Press has chosen to charge £5 for the volume, although it was advertised at 70s ... There are irritating errors ... Also, the translation is at times sadly deficient ... These shortcomings do not, however, seriously mar the value of the volume to the specialist.

From *Nature* 16 September 1961

100 Years Ago

On Saturday, September 9, the aërial post was inaugurated by Mr. Gustav Hamel, one of our most brilliant flyers, who carried a sack of letters in a Blériot monoplane from Hendon to Windsor in thirteen minutes ... The other aviators who should have started were prevented by the thirty-mile wind, and no further deliveries took place until Monday, when Messrs. Greswell and Driver carried six mail-bags over in the early morning ... The affair has aroused great interest, so much so that it is as well to sound a word of warning and say that the aëroplane post is neither practical, useful, nor economical. Letters can be sent far more cheaply, trustworthily, and conveniently by train or motor-van, and it is to be expected that these conditions will continue for the next half-century at least ... Besides, it is unthinkable for very many years to come that we should put good aviators to the menial task of carrying mails regularly.

From *Nature* 14 September 1911

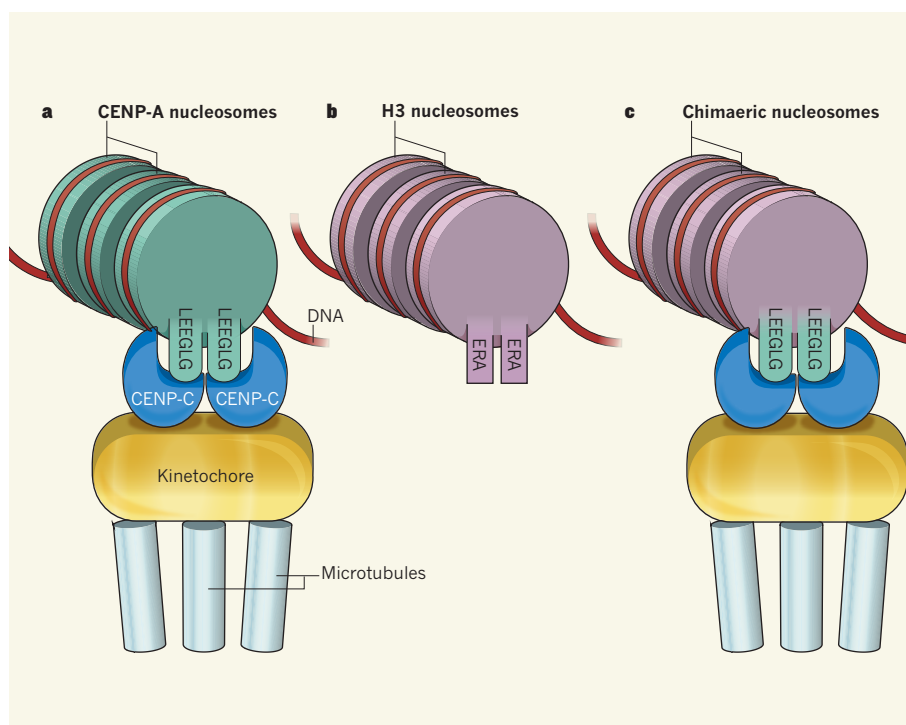


Figure 1 | Recognition of CENP-A-containing nucleosomes. **a**, Guse *et al.*¹ show that arrays of CENP-A nucleosomes allow active kinetochore-like structures that bind microtubules in frog egg extracts to assemble *in vitro*. This interaction is mediated by the centromeric protein CENP-C and requires the -LEEGLG amino-acid sequence at the carboxy terminus of CENP-A. **b**, H3 nucleosomes, which lack this particular motif and instead have three other residues (-ERA), do not assemble synthetic kinetochores. **c**, Replacement of -ERA at the carboxy terminus of H3 with -LEEGLG allows binding of CENP-C and assembly of synthetic kinetochores.

How do CENP-A nucleosomes differ from their H3 counterparts? Previous biophysical analyses hinted that the composition of CENP-A and H3 nucleosomes was radically different⁷. They suggested that CENP-A nucleosomes contain only a single subunit of H2A, H2B, CENP-A and H4, forming 'hemisomes' with half the height of typical nucleosomes.

In July, Tashiwana *et al.*² presented a high-resolution crystal structure of the human CENP-A nucleosome assembled *in vitro* on its natural substrate, α -satellite DNA. The structure shows that, like their H3 counterparts, *in vitro*-assembled CENP-A nucleosomes consist of eight subunits — with a CENP-A–H4 tetramer forming the core and bounded by two H2A–H2B dimers. But a clear difference is that CENP-A has a shorter helical domain at its amino terminus than H3. This alters DNA interactions at the nucleosome entry and exit points, meaning that CENP-A nucleosomes associate with a shorter stretch of DNA (121 compared with 147 base pairs). Moreover, two other amino-acid residues in one region (loop 1) cause it to jut out more in CENP-A nucleosomes, and their deletion affects CENP-A retention at, but not its targeting to, centromeres.

What attracts kinetochore proteins to CENP-A nucleosomes but not to H3 nucleosomes? Most differences between the properties

of CENP-A and H3 *in vivo* have been ascribed to a structural domain of CENP-A called CATD. This domain of CENP-A shows 22 differences in amino-acid residues compared with the same domain in H3 across a region of 40 residues. So, when placed within H3, it allows the resulting CENP-A–H3 chimaeric protein to be incorporated at centromeres⁴.

In Guse and colleagues' *in vitro* system¹, the CATD domain of CENP-A–H3 chimaeric proteins in nucleosome arrays was not sufficient or necessary for kinetochore assembly. Remarkably, when the authors replaced only the three carboxy-terminal residues of H3 (-ERA) with the six from CENP-A (-LEEGLG), this sequence allowed H3 to assemble kinetochore structures *in vitro* (Fig. 1). Further investigation showed that another centromeric protein, CENP-C, binds the -LEEGLG motif of human CENP-A — but only in the context of a nucleosome — to provide the platform for kinetochore formation. Presumably, once established, recognition of CENP-A nucleosomes, and their retention and replenishment at centromeres through subsequent cell divisions, are dictated by interactions with other kinetochore-associated proteins that may recognize additional differences such as loop 1 and CATD.

The *in vitro* formation of 'active' kinetochore complexes on purely CENP-A-containing nucleosomes¹ indicates that no neighbouring

H3 nucleosomes are required. H3 nucleosomes may, however, promote further stabilizing interactions between other kinetochore proteins and microtubules *in vivo*⁸. The conclusion that a key role of CENP-A chromatin is to provide a platform for kinetochore assembly is underscored by recent observations^{6,9,10} that the requirement for CENP-A for functional kinetochore formation can be bypassed by tethering the CENP-A chaperone protein HJURP — or other kinetochore proteins — to arrays of DNA binding sites.

The synthetic kinetochore-formation system described here¹ therefore provides

a new setting in which to further dissect the interactions required for the assembly of kinetochores on CENP-A chromatin. It will also allow investigation of the changes that occur in the composition of centromeric chromatin and kinetochores in response to cell-cycle events. ■

Alison Pidoux and Robin Allshire are at the Wellcome Trust Centre for Cell Biology, Institute of Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JR, UK.
e-mail: robin.allshire@ed.ac.uk

1. Guse, A., Carroll, C. W., Moree, B., Fuller, C. J. & Straight, A. F. *Nature* **477**, 354–358 (2011).
2. Tachibana, H. *et al. Nature* **476**, 232–235 (2011).
3. Allshire, R. C. & Karpen, G. H. *Nature Rev. Genet.* **9**, 923–937 (2008).
4. Black, B. E. & Cleveland, D. W. *Cell* **144**, 471–479 (2011).
5. Olszak, A. M. *et al. Nature Cell Biol.* **13**, 799–808 (2011).
6. Gascoigne, K. E. *et al. Cell* **145**, 410–422 (2011).
7. Dalal, Y., Wang, H., Lindsay, S. & Henikoff, S. *PLoS Biol.* **5**, e218 (2007).
8. Screpanti, E. *et al. Curr. Biol.* **21**, 391–398 (2011).
9. Kiermaier, E. *et al. Nature Cell Biol.* **11**, 1109–1115 (2009).
10. Barnhart, M. C. *et al. J. Cell Biol.* **194**, 229–243 (2011).

EARTH SCIENCE

Lethal volcanism

Data from the Siberian Traps volcanic region suggest that its magma source includes a significant component of recycled oceanic crust. This finding helps to explain why basalt eruptions are so environmentally devastating. SEE LETTER P.312

PAUL B. WIGNALL

Large igneous provinces (LIPs) are the most spectacular manifestation of volcanism on Earth. They consist of huge individual basaltic lava flows, with volumes measured in thousands of cubic kilometres, stacked layer upon layer to form vast volcanic plateaux. No one has ever witnessed a LIP eruption, which is probably just as well because geologists suspect that they cause mass extinctions and major perturbations to the carbon cycle. However, until now our understanding of the link between LIPs and environmental change has been poor. A new hypothesis from the Sobolev brothers, Stephan and Alexander, and their colleagues¹ (page 312 of this issue) may provide the answer.

Gas emissions are the obvious culprit when it comes to looking for a volcanism-triggered global catastrophe. Sulphur dioxide (SO₂) and halogens both cause short-term cooling and acid rain, with carbon dioxide — another abundant volcanic gas — being the well-known greenhouse gas. Clearly, there is scope for volcanic gases to cause environmental damage, and we can start to evaluate this potential by calculating the volumes emitted during a LIP eruption. This is straightforward: first measure how much gas is released during a modern basalt-lava eruption (in Hawaii, for example) and then simply scale up the figures to the size of a LIP province, which is typically several million cubic kilometres.

This calculation produces some very large numbers, but these are values for total gas emissions spread over the lifetime of the province, which may be a million years². This point is important because the gases added to

the atmosphere are scrubbed out within a few years, leaving ample time for recovery between eruptions. Even CO₂ is significantly drawn down within a few thousand years, consumed during rock weathering, especially weathering of the basalt flows themselves³.

Because of this rapid atmospheric recovery time, environmental damage must be done either by individual eruptions or by a series

of closely spaced ones. This assertion is supported by recent studies⁴ revealing that the onset of LIP eruptions coincides precisely with extinction events — the first punch does all the damage. But there is a problem. The volumes of gas emitted per individual lava flow are not especially impressive; they are less than the yearly anthropogenic pollution fluxes of gases such as SO₂ and CO₂ (ref. 2). So, here is the rub: LIP eruptions coincide with environmental crises and yet simple calculation of gas fluxes suggests that their impact should be minor.

Sobolev and colleagues' hypothesis¹ concerns the source of LIP magma and proposes that it is much more gas-rich than is generally assumed, especially at the onset of eruptions. Traditional dogma has LIPs originating from deep-generated mantle plumes that tap primitive material which has not undergone previous

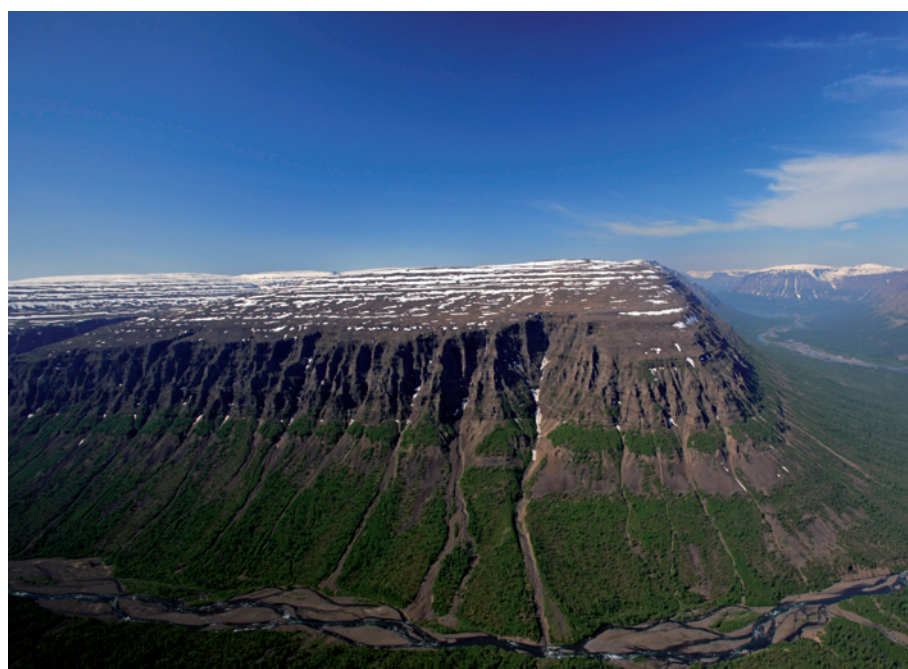


Figure 1 | Siberian flood-basalt flows in Putorana, Taymyr Peninsula. These rocks form part of a gigantic volcanic area that erupted 250 million years ago contemporaneously with the end-Permian mass extinction. A new study by Sobolev *et al.*¹ suggests that incorporation of large amounts of ocean-crust material in the erupting magma would have generated huge volumes of gases capable of triggering this environmental calamity.

melting⁵. Sobolev *et al.* do not disagree with this view, but they present geochemical and petrological evidence that ascending plumes incorporate oceanic crust that has been recycled into the mantle via ocean trenches, to the point where up to 20% of the plume may consist of this material. The consequences of this contamination are profound: it changes the way magma reaches Earth's surface and greatly increases the volume of gas released.

Traditionally, the arrival of a warm, buoyant mantle plume beneath the crust is thought to cause uplift and then stretching that allows magma to erupt⁶. A plume with a component of oceanic crust is considerably denser than a plume of mantle material alone, with the result that the process of ascent is a thermomechanical one: it 'eats' through the uppermost mantle and crust rather than rising by purely thermal processes. The thermomechanical erosion of lower crust requires more magma than is allowed by the traditional model and consequently a greater volume of gases (generated from the recycled oceanic crust). In Sobolev and colleagues' model¹, the volatile volcanic gases are driven off ahead of the basalt melting front, with the result that LIPs should start with a gigantic gaseous burp. The existence of a spectacular, initial explosive phase is supported by evidence from several LIPs, including the Siberian and Emeishan traps, which are closely linked with mass-extinction events⁴.

Sobolev and colleagues' model has much strength: it supplies the missing gas volumes, predicts the correlation with extinction and LIP eruption onset, and also explains the contentious lack of pre-eruption uplift (thermal doming) seen in many provinces⁷. Finally, the model provides an explanation for contemporaneous changes in carbon cycling recorded by the carbon isotope record. Mantle CO₂ has an isotopic composition that is not very different (it is slightly enriched in carbon-12) from that of the ocean-atmosphere system. This means that even eruption of huge volumes of volcanic CO₂ leaves little isotopic record. Despite this, most LIP eruptions, and especially those linked with mass extinctions, coincide with rapid, large swings of carbon isotope ratios that suggest large volumes of ¹²C-rich CO₂ are reaching the atmosphere. Carbon derived from oceanic crust is more ¹²C-rich than that from pure-mantle sources, and it may be this carbon that is leaving its signature in the carbon isotope record.

The work by Sobolev *et al.*¹ focuses on the Siberian Traps LIP (Fig. 1), which coincides with the great end-Permian extinction event around 250 million years ago. It will be instructive to test the model further by examining other LIPs, such as the Karoo-Ferrar Province, which coincides with only a minor extinction, and the Paraná-Etendeka Province, which had no impact on the global biota. Currently, various complex hypotheses⁸ are

invoked to explain the hit-and-miss environmental impact of LIPs. The missing variable in this relationship may be the degree to which plumes have reworked oceanic crust on their ascent to the surface. ■

Paul B. Wignall is at the School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK.
e-mail: p.wignall@see.leeds.ac.uk

ASTROPHYSICS

Rough times in the Galactic countryside

Knowledge of how the Milky Way formed and evolved is deficient. Simulations show that a past encounter with another galaxy may account for the Galaxy's intricate morphology. [SEE LETTER P.301](#)

CURTIS STRUCK

Our Local Group of galaxies, which includes the Milky Way, the Andromeda galaxy (M31) and their many smaller satellites, used to be viewed as a quiet, rural galactic neighbourhood. There was little evidence of significant galaxy collisions

or mergers in the group's history, except some involving the Magellanic Clouds. This is in rather stark contrast to some other nearby galaxy groups. The famous Antennae galaxies are violently rearranging each other, as they complete a merger between equals. The galaxies in the M81/M82 group are surrounded by large volumes of streams of interstellar gas



Figure 1 | Star stream. The extended stream of gas and stars that wraps around spiral galaxy NGC 5907, which is here seen edge-on, is thought to be the aftermath of a past encounter between the galaxy and a companion. The models of Purcell *et al.*² indicate that the Milky Way's Sagittarius stream, which is fainter and less massive than the stream of NGC 5907, is similarly the result of a past galaxy-galaxy interaction, and that this interaction may also account for the wave morphology in the Milky Way's disk.

melting⁵. Sobolev *et al.* do not disagree with this view, but they present geochemical and petrological evidence that ascending plumes incorporate oceanic crust that has been recycled into the mantle via ocean trenches, to the point where up to 20% of the plume may consist of this material. The consequences of this contamination are profound: it changes the way magma reaches Earth's surface and greatly increases the volume of gas released.

Traditionally, the arrival of a warm, buoyant mantle plume beneath the crust is thought to cause uplift and then stretching that allows magma to erupt⁶. A plume with a component of oceanic crust is considerably denser than a plume of mantle material alone, with the result that the process of ascent is a thermomechanical one: it 'eats' through the uppermost mantle and crust rather than rising by purely thermal processes. The thermomechanical erosion of lower crust requires more magma than is allowed by the traditional model and consequently a greater volume of gases (generated from the recycled oceanic crust). In Sobolev and colleagues' model¹, the volatile volcanic gases are driven off ahead of the basalt melting front, with the result that LIPs should start with a gigantic gaseous burp. The existence of a spectacular, initial explosive phase is supported by evidence from several LIPs, including the Siberian and Emeishan traps, which are closely linked with mass-extinction events⁴.

Sobolev and colleagues' model has much strength: it supplies the missing gas volumes, predicts the correlation with extinction and LIP eruption onset, and also explains the contentious lack of pre-eruption uplift (thermal doming) seen in many provinces⁷. Finally, the model provides an explanation for contemporaneous changes in carbon cycling recorded by the carbon isotope record. Mantle CO₂ has an isotopic composition that is not very different (it is slightly enriched in carbon-12) from that of the ocean-atmosphere system. This means that even eruption of huge volumes of volcanic CO₂ leaves little isotopic record. Despite this, most LIP eruptions, and especially those linked with mass extinctions, coincide with rapid, large swings of carbon isotope ratios that suggest large volumes of ¹²C-rich CO₂ are reaching the atmosphere. Carbon derived from oceanic crust is more ¹²C-rich than that from pure-mantle sources, and it may be this carbon that is leaving its signature in the carbon isotope record.

The work by Sobolev *et al.*¹ focuses on the Siberian Traps LIP (Fig. 1), which coincides with the great end-Permian extinction event around 250 million years ago. It will be instructive to test the model further by examining other LIPs, such as the Karoo-Ferrar Province, which coincides with only a minor extinction, and the Paraná-Etendeka Province, which had no impact on the global biota. Currently, various complex hypotheses⁸ are

invoked to explain the hit-and-miss environmental impact of LIPs. The missing variable in this relationship may be the degree to which plumes have reworked oceanic crust on their ascent to the surface. ■

Paul B. Wignall is at the School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK.
e-mail: p.wignall@see.leeds.ac.uk

ASTROPHYSICS

Rough times in the Galactic countryside

Knowledge of how the Milky Way formed and evolved is deficient. Simulations show that a past encounter with another galaxy may account for the Galaxy's intricate morphology. SEE LETTER P.301

CURTIS STRUCK

Our Local Group of galaxies, which includes the Milky Way, the Andromeda galaxy (M31) and their many smaller satellites, used to be viewed as a quiet, rural galactic neighbourhood. There was little evidence of significant galaxy collisions

or mergers in the group's history, except some involving the Magellanic Clouds. This is in rather stark contrast to some other nearby galaxy groups. The famous Antennae galaxies are violently rearranging each other, as they complete a merger between equals. The galaxies in the M81/M82 group are surrounded by large volumes of streams of interstellar gas



Figure 1 | Star stream. The extended stream of gas and stars that wraps around spiral galaxy NGC 5907, which is here seen edge-on, is thought to be the aftermath of a past encounter between the galaxy and a companion. The models of Purcell *et al.*² indicate that the Milky Way's Sagittarius stream, which is fainter and less massive than the stream of NGC 5907, is similarly the result of a past galaxy-galaxy interaction, and that this interaction may also account for the wave morphology in the Milky Way's disk.

and star clusters, which have been scattered by close encounters¹. But just at the time of year when it is nice to go out to a dark site and view the glorious and peaceful Milky Way, Purcell *et al.*² alert us to a prowler in the Galactic house — and evidence that it's been messing things up for some time. According to the authors' models, described on page 301 of this issue, the intruder may be responsible for the spiral-arm structure of the Milky Way, its central bar-shaped component and the flaring of its outermost disk.

The intruder is the Sagittarius Dwarf Elliptical Galaxy (SagDEG), also known as the Sgr Dwarf. SagDEG is not a very impressive object, and was only recognized³ as a galaxy in 1994. It is difficult to observe because it is near the disk plane on the opposite side of the Galactic Centre from the Solar System. It is also difficult to see because most of it has been ripped apart by the tidal forces of the Milky Way. The debris forms a huge, but very faint, star stream around the Milky Way (see Fig. 1 of the paper²). The observational evidence suggests that four or five tightly bound globular star clusters once orbited SagDEG. The cluster M54 may be its core. So the prowler seems to be a 'mouse'.

We see a mouse now, but Purcell and colleagues' models² indicate that a (shrinking) Galactic 'bear' did the damage to the disk of the Milky Way. Certainly, the visible galaxy would have been much more substantial before most of its stars were scattered. Moreover, the authors point to evidence, from observations and from models of cosmological-structure growth, that dwarf galaxies such as SagDEG's progenitor have massive haloes of dark matter — high mass-to-light ratios, in the jargon. Instead of consisting of a few globular clusters each a few million times the mass of the Sun, like SagDEG now, the progenitor may have been 100,000 times more massive.

The estimated size of the original galaxy is an extrapolation, and Purcell *et al.* consider a range of possibilities. The extent of the range is greater than that in some previous work^{4,5}, and will probably be a matter of contention in coming years. At the light end, the tidal effect of the galaxy on the Milky Way is reduced, although still significant. Nevertheless, as well as giving a possible history of the Milky Way, the results of this paper² are interesting as a demonstration of the possible effects of shrinking visitors in galactic households. There are various reasons to think that such effects could be widespread.

The first is that several dozen other dwarfs and star streams have been discovered around the Milky Way⁶ and Andromeda⁷, mostly in the past dozen years. With perhaps a couple of exceptions, these generally do not come nearly as close to the Milky Way as SagDEG does, and have not had the same kind of effects. But the progenitor of the Great Stream around Andromeda may have generated significant effects there^{8,9}. These dwarfs and streams have

very low surface brightness, so they would, in general, be difficult to detect in galaxies beyond the Local Group. A few have been detected in other galaxies (such as NGC 5907; Fig. 1), but presumably those extragalactic streams are among the brightest.

A second reason for considering that the effects are widespread is that, as noted in the paper, the effects of galaxy interactions may be long lived in some cases. A third reason is that high-resolution models of galaxy formation indicate¹⁰ that accretion onto galaxies out of the larger-scale structures that contain them continues throughout their cosmological history. The nature of this 'cold accretion' is not yet known. It may primarily consist of unformed streams of interstellar gas, or dwarf galaxies, or other constituents. The prowlers discovered in the Local Group suggest that at least some of these accreting objects are dwarf galaxies with dark haloes. These objects may therefore be common visitors and an important component of accretion onto the haloes of galaxies.

Because of the collective effects of dynamical friction, the relative orbits of most strongly interacting galaxies decay on timescales of hundreds of millions of years, and the end result is a merger of the two galaxies. Major mergers, which occur between two large galaxies of roughly equal mass, have been well studied, and we are getting an increasingly good understanding of their role in galaxy evolution.

Progress is also being made in understanding the role of minor mergers, which involve a companion a few to ten times smaller than the primary galaxy. Frequent stealthy invasions (and, ultimately, micro-mergers) by even smaller companions than those involved in minor mergers could generate waves in galaxy disks. This might, in turn, have long-term effects on disk evolution, giving us a new wrinkle in galaxy evolution. Beware of the wildlife, even in apparently quiet galaxies. ■

Curtis Struck is in the Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA.

e-mail: curt@iastate.edu

1. Struck, C. *Galaxy Collisions: Forging New Worlds from Cosmic Crashes* (Springer, 2011).
2. Purcell, C. W., Bullock, J. S., Tollerud, E. J., Rocha, M. & Chakrabarti, S. *Nature* **477**, 301–303 (2011).
3. Ibata, R. A., Gilmore, G. & Irwin, M. J. *Nature* **370**, 194–196 (1994).
4. Niederste-Ostholt, M., Belokurov, V., Evans, N. W. & Peñarrubia, J. *Astrophys. J.* **712**, 516–526 (2010).
5. Łokas, E. L. *et al.* *Astrophys. J.* **725**, 1516–1527 (2010).
6. Klement, R. J. *Astron. Astrophys. Rev.* **18**, 567–594 (2010).
7. Richardson, J. C. *et al.* *Astrophys. J.* **732**, 76 (2011).
8. Fardal, M. A., Babul, A., Guhathakurta, P., Gilbert, K. M. & Dodge, C. *Astrophys. J.* **682**, L33–L36 (2008).
9. Hammer, F. *et al.* *Astrophys. J.* **725**, 542–555 (2010).
10. Faucher-Giguère, C.-A., Keres, D. & Ma, C.-P. *Mon. Not. R. Astron. Soc.* (in the press); preprint at <http://arxiv.org/abs/1103.0001> (2011).

MULTIPLE SCLEROSIS

One protein, two healing properties

Multiple sclerosis is linked to rogue immune cells that attack mature neurons. Remarkably, immature neurons secrete a protein called LIF, which not only inhibits this attack, but also promotes repair of the damaged nerves.

SU M. METCALFE

Multiple sclerosis (MS) is a disabling autoimmune neurological disease that commonly affects young adults; in Britain alone there are more than 100,000 people with the disease. MS involves damage to the myelin sheath that normally insulates the electrical activity of nerve fibres. This in turn leads to a wide range of symptoms as specific nerves become inflamed and lose function. There is no cure. However, work on animal models has been encouraging, as it has shown that the transplantation of nerve progenitor cells not only inhibits the autoimmune attack that drives the disease, but also promotes the repair of damaged neurons¹. In fact, in

North America, human stem-cell transplantation is commercially available to patients with MS.

But is cell transplantation really necessary? Not according to Cao *et al.*², who report an exciting discovery in *Immunity*. They find that, at least in animal models of MS, a stem-cell-related cell-signalling protein called leukaemia inhibitory factor (LIF)³ can partially cure the disease. This finding opens the way for the development of a cell-free therapy for MS that is simple, safe and widely accessible.

Cao and colleagues studied mice that had experimental autoimmune encephalomyelitis (EAE) — a model of MS. They found that damage to the central nervous system was reduced not only by the intravenous delivery of neural

and star clusters, which have been scattered by close encounters¹. But just at the time of year when it is nice to go out to a dark site and view the glorious and peaceful Milky Way, Purcell *et al.*² alert us to a prowler in the Galactic house — and evidence that it's been messing things up for some time. According to the authors' models, described on page 301 of this issue, the intruder may be responsible for the spiral-arm structure of the Milky Way, its central bar-shaped component and the flaring of its outermost disk.

The intruder is the Sagittarius Dwarf Elliptical Galaxy (SagDEG), also known as the Sgr Dwarf. SagDEG is not a very impressive object, and was only recognized³ as a galaxy in 1994. It is difficult to observe because it is near the disk plane on the opposite side of the Galactic Centre from the Solar System. It is also difficult to see because most of it has been ripped apart by the tidal forces of the Milky Way. The debris forms a huge, but very faint, star stream around the Milky Way (see Fig. 1 of the paper²). The observational evidence suggests that four or five tightly bound globular star clusters once orbited SagDEG. The cluster M54 may be its core. So the prowler seems to be a 'mouse'.

We see a mouse now, but Purcell and colleagues' models² indicate that a (shrinking) Galactic 'bear' did the damage to the disk of the Milky Way. Certainly, the visible galaxy would have been much more substantial before most of its stars were scattered. Moreover, the authors point to evidence, from observations and from models of cosmological-structure growth, that dwarf galaxies such as SagDEG's progenitor have massive haloes of dark matter — high mass-to-light ratios, in the jargon. Instead of consisting of a few globular clusters each a few million times the mass of the Sun, like SagDEG now, the progenitor may have been 100,000 times more massive.

The estimated size of the original galaxy is an extrapolation, and Purcell *et al.* consider a range of possibilities. The extent of the range is greater than that in some previous work^{4,5}, and will probably be a matter of contention in coming years. At the light end, the tidal effect of the galaxy on the Milky Way is reduced, although still significant. Nevertheless, as well as giving a possible history of the Milky Way, the results of this paper² are interesting as a demonstration of the possible effects of shrinking visitors in galactic households. There are various reasons to think that such effects could be widespread.

The first is that several dozen other dwarfs and star streams have been discovered around the Milky Way⁶ and Andromeda⁷, mostly in the past dozen years. With perhaps a couple of exceptions, these generally do not come nearly as close to the Milky Way as SagDEG does, and have not had the same kind of effects. But the progenitor of the Great Stream around Andromeda may have generated significant effects there^{8,9}. These dwarfs and streams have

very low surface brightness, so they would, in general, be difficult to detect in galaxies beyond the Local Group. A few have been detected in other galaxies (such as NGC 5907; Fig. 1), but presumably those extragalactic streams are among the brightest.

A second reason for considering that the effects are widespread is that, as noted in the paper, the effects of galaxy interactions may be long lived in some cases. A third reason is that high-resolution models of galaxy formation indicate¹⁰ that accretion onto galaxies out of the larger-scale structures that contain them continues throughout their cosmological history. The nature of this 'cold accretion' is not yet known. It may primarily consist of unformed streams of interstellar gas, or dwarf galaxies, or other constituents. The prowlers discovered in the Local Group suggest that at least some of these accreting objects are dwarf galaxies with dark haloes. These objects may therefore be common visitors and an important component of accretion onto the haloes of galaxies.

Because of the collective effects of dynamical friction, the relative orbits of most strongly interacting galaxies decay on timescales of hundreds of millions of years, and the end result is a merger of the two galaxies. Major mergers, which occur between two large galaxies of roughly equal mass, have been well studied, and we are getting an increasingly good understanding of their role in galaxy evolution.

Progress is also being made in understanding the role of minor mergers, which involve a companion a few to ten times smaller than the primary galaxy. Frequent stealthy invasions (and, ultimately, micro-mergers) by even smaller companions than those involved in minor mergers could generate waves in galaxy disks. This might, in turn, have long-term effects on disk evolution, giving us a new wrinkle in galaxy evolution. Beware of the wildlife, even in apparently quiet galaxies. ■

Curtis Struck is in the Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA.

e-mail: curt@iastate.edu

1. Struck, C. *Galaxy Collisions: Forging New Worlds from Cosmic Crashes* (Springer, 2011).
2. Purcell, C. W., Bullock, J. S., Tollerud, E. J., Rocha, M. & Chakrabarti, S. *Nature* **477**, 301–303 (2011).
3. Ibata, R. A., Gilmore, G. & Irwin, M. J. *Nature* **370**, 194–196 (1994).
4. Niederste-Ostholt, M., Belokurov, V., Evans, N. W. & Peñarrubia, J. *Astrophys. J.* **712**, 516–526 (2010).
5. Łokas, E. L. *et al.* *Astrophys. J.* **725**, 1516–1527 (2010).
6. Klement, R. J. *Astron. Astrophys. Rev.* **18**, 567–594 (2010).
7. Richardson, J. C. *et al.* *Astrophys. J.* **732**, 76 (2011).
8. Fardal, M. A., Babul, A., Guhathakurta, P., Gilbert, K. M. & Dodge, C. *Astrophys. J.* **682**, L33–L36 (2008).
9. Hammer, F. *et al.* *Astrophys. J.* **725**, 542–555 (2010).
10. Faucher-Giguère, C.-A., Keres, D. & Ma, C.-P. *Mon. Not. R. Astron. Soc.* (in the press); preprint at <http://arxiv.org/abs/1103.0001> (2011).

MULTIPLE SCLEROSIS

One protein, two healing properties

Multiple sclerosis is linked to rogue immune cells that attack mature neurons. Remarkably, immature neurons secrete a protein called LIF, which not only inhibits this attack, but also promotes repair of the damaged nerves.

SU M. METCALFE

Multiple sclerosis (MS) is a disabling autoimmune neurological disease that commonly affects young adults; in Britain alone there are more than 100,000 people with the disease. MS involves damage to the myelin sheath that normally insulates the electrical activity of nerve fibres. This in turn leads to a wide range of symptoms as specific nerves become inflamed and lose function. There is no cure. However, work on animal models has been encouraging, as it has shown that the transplantation of nerve progenitor cells not only inhibits the autoimmune attack that drives the disease, but also promotes the repair of damaged neurons¹. In fact, in

North America, human stem-cell transplantation is commercially available to patients with MS.

But is cell transplantation really necessary? Not according to Cao *et al.*², who report an exciting discovery in *Immunity*. They find that, at least in animal models of MS, a stem-cell-related cell-signalling protein called leukaemia inhibitory factor (LIF)³ can partially cure the disease. This finding opens the way for the development of a cell-free therapy for MS that is simple, safe and widely accessible.

Cao and colleagues studied mice that had experimental autoimmune encephalomyelitis (EAE) — a model of MS. They found that damage to the central nervous system was reduced not only by the intravenous delivery of neural

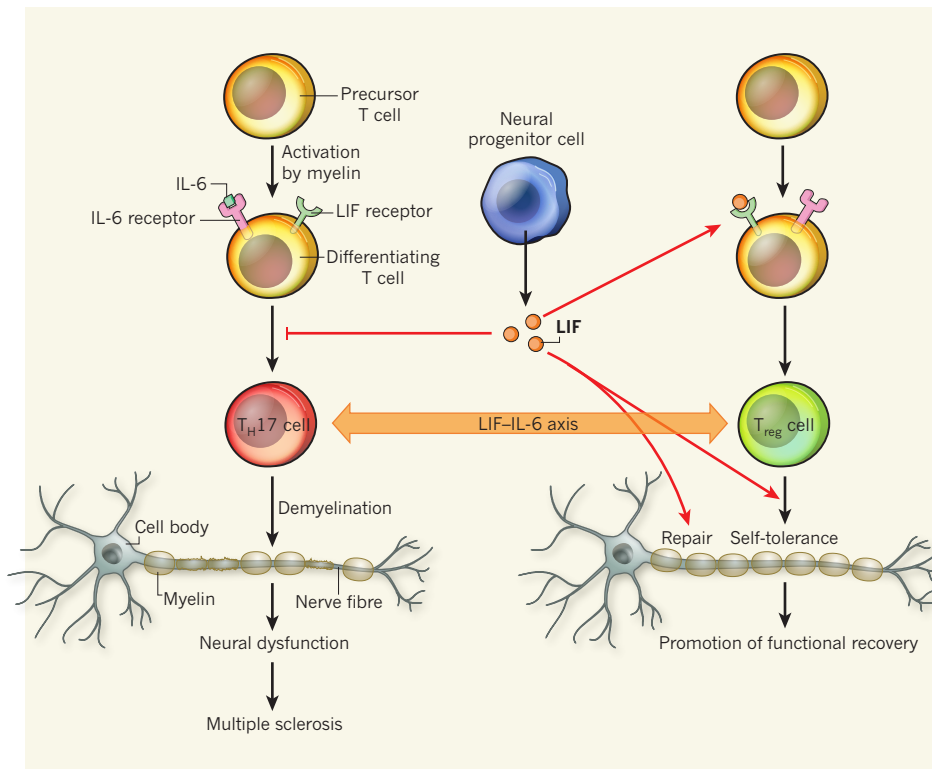


Figure 1 | Multiple sclerosis and treatment with LIF. In multiple sclerosis, interaction between IL-6 and its receptor on the surface of precursor T cells in the central nervous system drives differentiation of TH17 cells. These pathogenic cells cause autoimmune inflammatory damage to the myelin sheath surrounding nerve fibres. Cao *et al.*² show that neural progenitor cells release LIF, which selectively inhibits the differentiation of TH17 cells by opposing IL-6-mediated signalling pathways. The mechanism underlying this inhibition involves a LIF–IL-6 axis that operates a cell-fate control switch in T-cell differentiation, with the gp190 subunit on the LIF receptor being pivotal⁵. In addition to inhibiting TH17-cell differentiation, LIF promotes differentiation of protective Treg cells, supports Treg-cell-mediated ‘self-tolerance’ and can directly aid repair in the central nervous system. Therefore, there is a strong case for the use of LIF in treating multiple sclerosis (red arrows).

progenitor cells, but also by simply injecting the medium in which these cells were cultured. This suggested that a cell-derived soluble factor provides the protective effect, and the authors’ screening studies revealed that the crucial factor is LIF. Indeed, commercially available LIF alone successfully replaced the cell therapy².

Cao *et al.* next turned their efforts to investigating exactly how LIF exerts its beneficial effect. They found that it acts through suppression of a specific type of immune cell called a TH17 cell⁴. This class of T cell functions to defend the gut and mucosal tissues from invading pathogens. Sometimes, however, rogue TH17 cells arise within otherwise healthy host tissues, leading to autoimmune inflammation — as occurs in the central nervous system of patients with MS.

Notably, a mediator of inflammation called IL-6 is essential for TH17-cell development; and herein lies a twist in the tale. Structurally, IL-6 and LIF are very closely related. So how is it that IL-6 exacerbates MS, whereas LIF protects against the disease? The answer lies in the balance between LIF and IL-6 in affecting T-cell differentiation⁵ — the ‘LIF–IL-6 axis’.

This pivotal axis is a cell-fate decision fork leading to either an IL-6-driven TH17-cell lineage or a LIF-driven Treg-cell lineage. Notably, the Treg-cell lineage is protective, promoting self-tolerance.

To respond to LIF or IL-6, cells must express the corresponding receptors at their surface. Of the two subunits of the LIF receptor (gp190 and gp130), gp190 endows specificity for LIF binding. In the IL-6 receptor, however, both subunits are gp130, and so LIF cannot activate it. Cao *et al.* show that undifferentiated peripheral T cells obtained from either EAE mice or patients with MS undergo transient expression of gp190 when activated. When LIF was added, the cells retained gp190 and so — through activation of an inhibitory signalling cascade, the ERK pathway — failed to mature into the inflammatory, myelin-attacking TH17 cells. Thus, in EAE and MS, a yin–yang type of LIF versus IL-6 regulation seems to operate, which is determined by the expression of gp190 (Fig. 1).

The fate of the immature, LIF-blocked TH17 cells remains ambiguous. The authors find no evidence that LIF causes these cells to mature down another lineage². However, previous

work⁵ suggests that the cells would develop into Treg cells in response to LIF. This uncertainty ought to be resolved because, if Treg cells indeed arise, the myelin-protective immunity offered by these cells would perpetuate the beneficial effect of LIF therapy.

By identifying LIF as a potential treatment for MS, the present paper holds promise of early translation to the clinic. However, soluble LIF cannot be used for treatment purposes because it is rapidly degraded by protease enzymes in the blood. To overcome this problem, my team and our collaborators have developed LIF in the form of biodegradable nanoparticles^{5,6}. This formulation was successful not only in providing a slow-release vehicle for LIF, but also in operating as a ‘magic bullet’ to target LIF directly to specific cell types — for example, to T cells for the induction of Treg cells, or to nerve cells for their repair. Another reported⁷ therapeutic approach involves delivering LIF by means of a viral vector.

But can LIF be beneficial in patients with MS who already harbour fully differentiated inflammatory TH17 cells? A recent clinical trial⁸ is highly relevant to answering this question. The trial involved depletion of circulating T cells — including TH17 cells — using the therapeutic antibody alemtuzumab. T-cell populations were then allowed to recover, from undifferentiated precursors resistant to alemtuzumab. The newly emerging populations included Treg-type cells that secreted neuroprotective factors in response to products of damaged myelin; this cell population was absent from pretreatment blood samples taken from the patients. The beneficial effects of the treatment are profound, continuing over several years. These effects are consistent with the idea that newly arising, myelin-reactive Treg-type cells form a self-sustaining population of mature cells that can oppose the maturation of myelin-reactive TH17 cells.

Clearly, great strides towards improved treatment of MS are being made. New means of exploiting the natural protective properties of LIF in both the immune and the nervous systems are now available to take Cao and colleagues’ results into preclinical studies. Future studies in which T-cell depletion is combined with LIF therapy are eagerly awaited. ■

Su M. Metcalfe is in the Brain Repair Centre, Department of Neurology, University of Cambridge, Addenbrooke’s Hospital, Cambridge CB2 0PY, UK.
e-mail: smm1001@cam.ac.uk

1. Pluchino, S. *et al.* *Nature* **422**, 688–694 (2003).
2. Cao, W. *et al.* *Immunity* **35**, 1–12 (2011).
3. Metcalfe, D. *Stem Cells* **21**, 5–14 (2003).
4. Harrington, L. E. *et al.* *Nature Immunol.* **6**, 1123–1132 (2005).
5. Gao, W. *et al.* *Cell Cycle* **8**, 1444–1450 (2009).
6. Park, J. *et al.* *Mol. Pharmacol.* **8**, 143–152 (2011).
7. Slaets, H. *et al.* *Mol. Ther.* **18**, 684–691 (2010).
8. Jones, J. L. *et al.* *Brain* **133**, 2232–2247 (2010).

Mouse genomic variation and its effect on phenotypes and gene regulation

Thomas M. Keane^{1*}, Leo Goodstadt^{2*}, Petr Danecek^{1*}, Michael A. White³, Kim Wong¹, Binnaz Yalcin², Andreas Heger⁴, Avigail Agam^{2,4}, Guy Slater¹, Martin Goodson², Nicholas A. Furlotte⁵, Eleazar Eskin⁵, Christoffer Nellåker⁴, Helen Whitley², James Cleak², Deborah Janowitz^{2,6}, Polinka Hernandez-Pliego², Andrew Edwards², T. Grant Belgard⁴, Peter L. Oliver⁴, Rebecca E. McIntyre¹, Amarjit Bhomra², Jérôme Nicod², Xiangchao Gan², Wei Yuan², Louise van der Weyden¹, Charles A. Steward¹, Sendu Bala¹, Jim Stalker¹, Richard Mott², Richard Durbin¹, Ian J. Jackson⁷, Anne Czechanski⁸, José Afonso Guerra-Assunção⁹, Leah Rae Donahue⁸, Laura G. Reinholdt⁸, Bret A. Payseur³, Chris P. Ponting⁴, Ewan Birney⁹, Jonathan Flint² & David J. Adams¹

We report genome sequences of 17 inbred strains of laboratory mice and identify almost ten times more variants than previously known. We use these genomes to explore the phylogenetic history of the laboratory mouse and to examine the functional consequences of allele-specific variation on transcript abundance, revealing that at least 12% of transcripts show a significant tissue-specific expression bias. By identifying candidate functional variants at 718 quantitative trait loci we show that the molecular nature of functional variants and their position relative to genes vary according to the effect size of the locus. These sequences provide a starting point for a new era in the functional analysis of a key model organism.

Until the end of the 20th century the molecular basis for morphological, physiological, biochemical and behavioural variation in laboratory mice remained largely obscure^{1–3}. At the beginning of the 21st century, decoding the complete genome of one strain, C57BL/6J, the mouse reference genome, revolutionized our ability to relate sequence to function^{4,5}. It enabled genetic screens in mice to be performed on an unprecedented scale⁶, it facilitated the task of creating a complete set of null alleles for all genes^{7,8}, and it accelerated the discovery of mouse sequence diversity^{9,10}.

Our catalogues, however, remain incomplete and some forms of variation are largely undocumented. Whereas we now know more about the extent of phenotypic variation among laboratory strains of mice^{11–16} and the complexity of genetic action, from fully penetrant Mendelian effects, partially penetrant modifiers^{17,18} and non-additive effects¹⁸, to the quasi-infinitesimal genetic architecture that underlies most quantitative traits¹⁹, we are still largely ignorant of the molecular basis of the majority of genetically influenced phenotypes.

Here we describe the generation and analysis of sequence from 17 key mouse genomes, obtained using next-generation sequencing^{20,21}. The genomes include those of the classical laboratory strains C3H/HeJ, CBA/J, A/J, AKR/J, DBA/2J, LP/J, BALB/cJ, NZO/HILtJ and NOD/ShiLtJ, and those of four wild-derived inbred strains CAST/EiJ, PWK/PhJ, WSB/EiJ and SPRET/EiJ, which include the progenitors of the common laboratory strains and are representative of the *Mus musculus castaneus*, *Mus musculus musculus*, *Mus musculus domesticus* and *Mus spretus* taxa, respectively. We also sequenced three related 129-strains, (129S5SvEv^{Brd}, 129P2/OlaHsd and 129S1/SvImJ) representing the genetic backgrounds on which more than 5,000 knockout mice have been generated²² and C57BL/6NJ, the strain used by the genome-wide knockout programmes KOMP, NorCOMM and EUCOMM^{7,8,22}. Collectively the sequences of these strains capture

the genomes of most of the commonly used strains of mice and their progenitors^{14,23–25}.

We document the variation we have discovered, describe the distribution of variants between strains, and explore the evolutionary origins of the subspecies that gave rise to the laboratory mouse. Using two examples we demonstrate how the sequence can be used to investigate the molecular origins of phenotypic variation. First, we use sequence variation to assay allele-specific variation in gene expression. We show how, in combination with a measure of activity at gene promoters, it is possible to implicate functional variants in gene expression regulation. Second, we explore the molecular basis of quantitative traits. We ask whether functional variants responsible for quantitative variation have common molecular features, in terms of their position (inside or outside genes) and their molecular class (single nucleotide polymorphisms (SNPs), indels or structural variants).

Data generation and variant discovery

Figure 1 and Table 1 summarize the sequence generated and the variants discovered. We defined all sequence as either the same as, or different from, that of the reference strain (C57BL/6J; MGSCv37 assembly) and we report our results with respect to an accessible genome: those sites to which sequence reads can be uniquely mapped with mapping qualities greater than 40 (Supplementary Methods). This represented on average 83.8% of the reference genome and 94.7% of coding sequence of each strain.

Between 13% and 23% of each genome is inaccessible (Table 1 and Supplementary Figs 1–17). The higher proportion of inaccessible regions in the wild-derived strains indicates that divergence from the mouse reference is a major contributor to inaccessibility. In the accessible mouse genome, we identified 56.7 million (M) unique

¹The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1HH, UK. ²The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. ³Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA. ⁴MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3QX, UK. ⁵University of California, Los Angeles, California 90095-1596, USA. ⁶Department of Psychiatry and Psychotherapy, Ernst-Moritz-Arndt-Universität Greifswald Klinikum der Hansestadt Stralsund, Rostocker Chaussee 70, 18437 Stralsund, Germany. ⁷Medical Research Council Human Genetics Unit, Crewe Road, Edinburgh EH4 2XU, UK. ⁸The Jackson Laboratory, Bar Harbor, Maine 04609, USA. ⁹European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK.

*These authors contributed equally to this work.

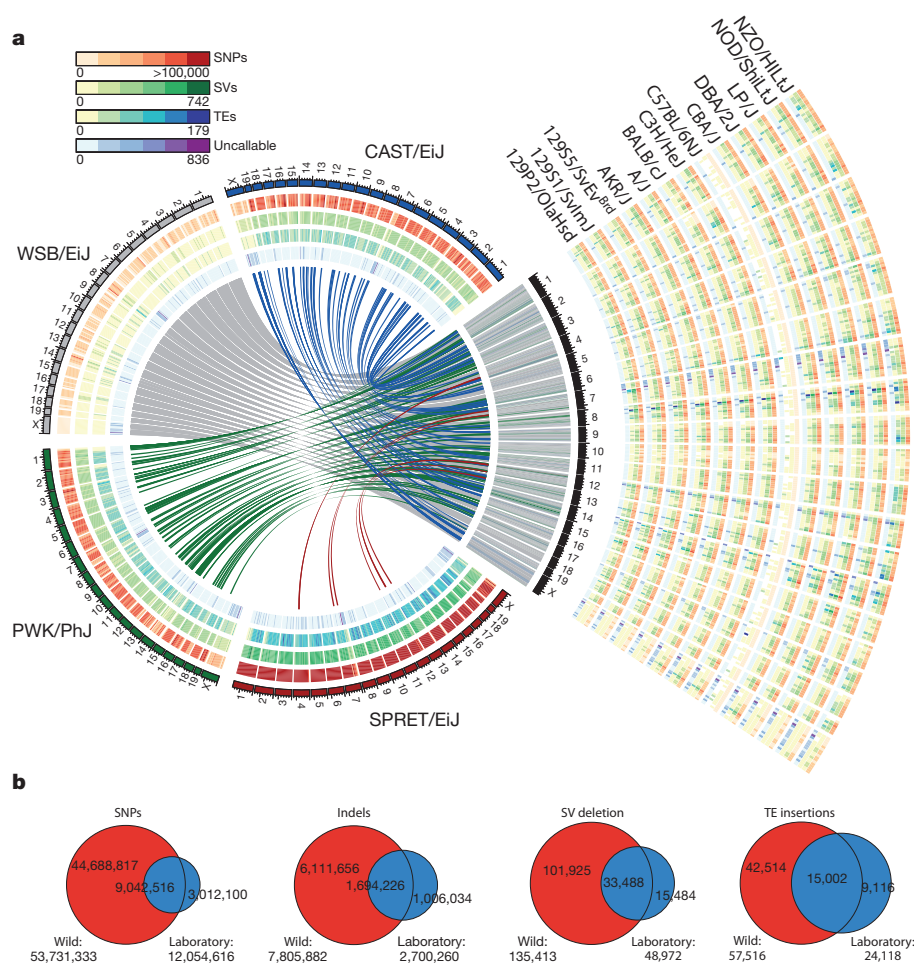


Figure 1 | An overview of variants called from 17 mouse genomes relative to the reference. **a**, The four wild-derived strains (CAST/EiJ, WSB/EiJ, PWK/PhJ and SPRET/EiJ) are representative of the *Mus musculus castaneus*, *Mus musculus musculus*, *Mus musculus domesticus* and *Mus spretus* taxa and include the progenitors from which the classical laboratory strains were derived. These genomes are shown in a circle with tracks indicating the relative density of SNPs, structural variants (SVs) and uncallable regions (binned into 10-Mb regions). Transposable element (TE) insertions, which are a subset of the

structural variant calls, are shown as a separate track. Corresponding tracks are shown for each of the 13 classical laboratory strains to the right of the circle. Links crossing the circle indicate regions on the reference where the wild-derived strain is closest to the reference (375-kb bins). **b**, The numbers inside the Venn diagrams indicate the number of SNPs, indels, structural variant deletions and transposable element insertions in the wild-derived and classical laboratory strains. The numbers beneath each Venn diagram indicate totals for each type of variant in the wild and classical laboratory strains.

SNPs, 8.8M unique indels and 0.28M structural variants including 0.07M transposable element insertion sites (Table 1).

The sensitivity and specificity of our variant calls were established using 17.5 million bases (Mb) of DNA from one non-reference strain (NOD/ShiLtJ) that we generated with established sequencing technology. We sequenced 107 bacterial artificial chromosomes (BACs)²⁶ spread over loci on chromosomes 1, 6, 11 and 17. The sequence has an estimated accuracy of one error per 100,000 base pairs (bp). We aligned 16.2 Mb of the BAC sequence to the MGSCv37 mouse reference and from that estimated that 3.6% of our next-generation-derived NOD/ShiLtJ SNP calls were false positives, and 6.5% were false negatives. We compared our genotype calls to those in public databases and found over 99.4% and 99.1% agreement with the two largest SNP data sets (Perlegen⁹ and dbSNP²⁷). However, we also found that these data sets have large false-negative rates of 83.7% and 84.1%, respectively.

We identified far fewer indels (1–100 bp) than SNPs and with lower confidence (Table 1). We relied for validation on comparison with the NOD/ShiLtJ BAC sequences and estimated false-positive and -negative rates to be 2.2% and 20.1%, respectively. Collectively, we estimate an average of 2.61 sequence errors per 10 kilobases (kb) of accessible sequence, an accuracy of 99.97% in NOD/ShiLtJ, which should extend to the other sequenced strains.

We used the NOD/ShiLtJ BAC sequence to estimate how many variants are contained within inaccessible regions. We found that the BAC sequence in inaccessible regions has approximately 2.8 times more SNPs per base than the rest of the BAC sequence. Sequence reads could not be unambiguously mapped to these regions, resulting in missed variant calls. An analysis of the content of the inaccessible sequence is provided in Supplementary Table 1. Our analysis of the NOD/ShiLtJ BAC sequence implies that at least 30% of all SNPs in the genomes of the strains we sequenced remain to be discovered. The majority of these SNPs are located in intergenic regions of the genome. In addition to homozygous SNP positions we also called 5.2M heterozygous positions. These result from misalignments around indels and structural variant breakpoints, duplicated loci and low depth positions.

We called 0.71M structural variants >100 bp (0.41M simple deletions, 0.29M simple insertions, 2,100 inversions, 1,556 copy number gains and 3,658 complex structural variants) (Table 1 and Fig. 1) at 0.28M positions, as described in our accompanying paper²⁸. On average 48.4 Mb of sequence of each strain falls into structurally variant regions of the genome (33 Mb for the laboratory strains and 98.2 Mb for wild-derived strains). Structural variants cluster with SNPs in each strain (Supplementary Fig. 1–17), indicating that the vast majority of these events may be ancestral in origin. This may also reflect high rates of polymorphism consequent to break-induced replication involved in

Table 1 | An overview of the sequence and variants called from 17 mouse genomes.

Strain	Gb of mapped data	Coverage	% of genome inaccessible	SNPs	(Private)	Indels	(Private)	Structural variants	(Private)
C57BL/6NJ	77.29	29.29	13.21	9,844	(1,488)	22,228	(4,259)	431	(75)
129S1/SvImJ	71.91	27.25	15.30	4,458,004	(1,489)	886,136	(16,140)	29,153	(786)
129S5SvEv ^{Brd}	50.27	19.05	15.17	4,383,799	(1,991)	810,310	(21,214)	25,340	(691)
129P2/Ola	115.52	43.78	14.47	4,694,529	(23,677)	1,028,629	(58,173)	32,227	(3,430)
A/J	70.39	26.68	15.90	4,198,324	(44,837)	823,688	(24,502)	28,691	(1,474)
AKR/J	107.16	40.61	14.86	4,331,384	(87,527)	966,002	(64,422)	30,742	(3,576)
BALB/cJ	65.72	24.90	15.09	3,920,925	(29,973)	831,193	(30,998)	25,702	(1,056)
C3H/HeJ	92.81	35.17	15.09	4,403,599	(16,804)	949,206	(34,834)	28,532	(1,779)
CBA/J	77.43	29.34	14.79	4,511,278	(34,203)	929,860	(35,976)	28,183	(1,178)
DBA/2J	65.11	24.67	15.09	4,468,071	(72,214)	868,611	(37,085)	28,346	(1,469)
LP/J	73.03	27.67	15.29	4,701,445	(53,509)	947,614	(33,817)	30,024	(1,194)
NOD/ShiLtJ	75.88	28.75	17.30	4,323,530	(143,489)	797,086	(41,113)	30,605	(2,479)
NZO/HILtJ	45.68	17.31	16.06	4,492,372	(210,256)	806,511	(60,231)	25,125	(1,938)
PWK/PhJ	66.99	25.38	19.26	17,202,436	(4,461,772)	2,635,885	(833,794)	90,125	(25,383)
CAST/EiJ	64.84	24.57	19.18	17,673,726	(5,368,019)	2,727,089	(956,828)	86,322	(25,232)
WSB/EiJ	48.19	18.26	16.23	6,045,573	(894,875)	1,197,006	(211,348)	35,066	(5,957)
SPRET/EiJ	70.41	26.68	23.26	35,441,735	(23,455,525)	4,456,243	(2,936,998)	157,306	(91,721)
Total	1,238.63	469.36		129,260,574		21,683,297		711,920	

Private variants are strain-specific variants.

the production of a structural variant²⁹. Only 7.5% of structural variants were private to one of the classical laboratory strains.

Functional variants

We identified 0.12M SNP positions in protein-coding sequence that lead to amino acid changes (non-synonymous substitutions) and 0.26M that do not (synonymous substitutions). In total 2,051 stop codons across all strains and transcripts were discovered, an average of 85 for the classical laboratory strains and 251 for the wild-derived strains. Supplementary Fig. 18 shows the distribution of these variants across the strains. Non-synonymous changes are seen, on average, every 1,454 codons, and rarely cluster. Extreme variation, however, occurs within a coding exon of *Prdm9*, a 'speciation gene'³⁰, whose zinc-finger-encoding domains vary greatly across the sequenced strains (Supplementary Fig. 19). By sequencing RNA we confirmed 99.84% of the coding SNPs that were covered by 10 or more RNA-Seq reads in expressed genes (Supplementary Table 2).

Some functional variants previously reported in one strain were found for the first time in others. In LP/J mice we identified a mutation in the DNA polymerase ϵ (*Poli*) gene. This premature stop codon, which ablates gene function, has previously been identified in 129-derived mice (MMU18:70688442)³¹. We also discovered that a mutation in *Disc1*, known in 129-derived mice and associated with a deficit in working memory³², is also present in LP/J. Further, we discovered a truncating mutation (MMU10:53345838) in the mini-chromosome maintenance gene *Mcm9* (ref. 33) in SPRET/EiJ. This gene is thought to have an important role in replication, suggesting functional redundancy or the existence of a paralogous gene in SPRET/EiJ.

Variation between mouse strains

The classical laboratory strains of mice carried relatively few private variants (~2% of all variants called in each strain) (Table 1). These variants were distributed genome-wide, indicating that they had either arisen since the divergence of these strains (Supplementary Fig. 1–17), or are errors. We observed significant differences in transposable element families across the laboratory and wild-derived strains (Fig. 1). Transposon element variants (TEVs) were found to be depleted near transcriptional start sites, in or near exons, and long interspersed nuclear element (LINE) variants were depleted within the introns of transcription factor genes. Within introns, we find a significantly reduced number of endogenous retroviral (ERV) TEVs that are inserted in the sense transcriptional orientation.

Loci that are absent from the C57BL/6J reference genome are difficult to access. We identified 424 Mb of novel sequence (contigs

>100 bp; 48.4 Mb for contigs >1 kb)(Supplementary Fig. 20). Unsurprisingly, more is found in the wild-derived strains than in the classical laboratory strains, which are largely derived from a common pool of founders. Of the novel sequence 20.4 Mb aligned with the Celera mixed strain assembly³⁴ and other mouse sequence not present in the reference genome; 562.9 kb mapped to the rat reference genome and 18.9 kb to the rabbit reference. About 30 Mb of novel sequence was conserved across multiple strains (Supplementary Fig. 20).

The phylogenetic history of the mouse

We used the accessible sequences of the wild-derived strains to explore the evolutionary history of the primary subspecies that gave rise to the laboratory mouse. We conducted a Bayesian concordance analysis³⁵ with the sequences of *M. m. musculus* (PWK/PhJ), *M. m. domesticus* (WSB/EiJ), *M. m. castaneus* (CAST/EiJ) and *M. spretus* (SPRET/EiJ), using rat as an outgroup.

We observed substantial phylogenetic discordance across the genomes of *M. m. musculus*, *M. m. domesticus* and *M. m. castaneus* (Fig. 2). In the face of this discordance, we identified a *M. m. musculus*/*M. m. castaneus* primary subspecies history (concordance factor (CF) = 37.9%; 95% credibility interval (CI) = 37.8–38.0%). The two other possible histories were supported by equal numbers of loci (CF = 30.3%; 95% CI = 30.2–30.4%; and CF = 30.2%; 95% CI = 30.1–30.3%), closely matching expectations from theoretical models of incomplete lineage sorting^{36–38}. Phylogenetic switching occurs over a short physical scale, in rough agreement with the spatial pattern of linkage disequilibrium in natural populations of house mice³⁹, and median locus sizes parallel the three phylogenetic histories (primary history, 40,975 bp; alternative histories, 33,626 bp and 33,412 bp). Despite its considerable divergence time from house mice, we also found phylogenetic discordance involving *M. spretus*: 12.1% of loci did not place this species as the outgroup to a *M. musculus* subspecies clade.

Allele-specific functional differences

We combined a measure of allele-specific variation with a measure of activity at gene promoters to implicate functional variants. Sequencing RNA from an F1 hybrid of two sequenced strains and assaying the relative abundance of allelic variants in transcripts makes it possible to assess the variation in gene expression. We sequenced RNA from six tissues (liver, thymus, spleen, lung, hippocampus and heart (Supplementary Table 2)) obtained from an F1 generated by crossing the reference strain (C57BL/6J) with one sequenced strain (DBA/2J). A total of 40,521 SNP positions were covered by RNA reads spread over

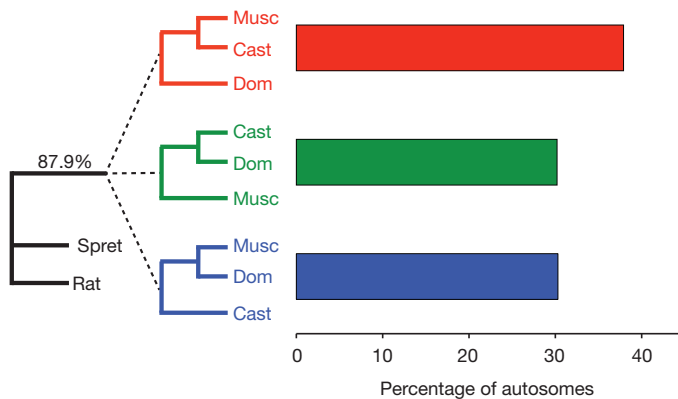


Figure 2 | Genomic partitioning of phylogenetic history. Bayesian concordance factors were estimated from 43,255 individual locus trees. 87.9% of loci place *M. spretus* (Spret) and rat as the outgroup to the *M. musculus* subspecies. Within *M. musculus*, there is a primary history supporting a *M. m. musculus* (Musc)/*M. m. castaneus* (Cast) sister relationship (37.9%) with *M. m. domesticus* (Dom) branching off first. The two alternative topologies are supported by equal percentages of the genome (30.3% and 30.2%). 95% credibility intervals on all estimates are $\pm 0.1\%$.

15,884 genes (≥ 1 read per gene), of which 6,975 had at least 20 reads crossing SNP positions⁴⁰.

We define allelic bias as the proportion of expression attributable to a particular parental strain, ranging from 0 to 1, with the null hypothesis of 0.5 in the absence of any bias. Due to the very high abundance of RNA sequence data and of SNPs within many genes revealed by whole genome sequencing, many (41%) loci show a significant bias towards one or other allele in at least one tissue; 12% of all loci showed a substantial expression bias (expression below 25% or above 75% of the reference allele).

Figure 3 shows the distribution of allele-specific biases between tissue pairs at the gene level and Supplementary Table 3 shows the concordance of allele-specific biases for each pair of tissues examined. 2,871 genes were found to be significantly different (0.01 false discovery rate, FDR) in at least one tissue pair (Supplementary Table 4). Most differences (95%) between tissues were due to biased allelic expression occurring in one tissue only. However, 336 (4.8%) of tested transcripts showed a different pattern: they show a biased allelic expression in more than one tissue, but the bias occurs in opposing directions. One example is the *Phb* gene: in liver, 76% of informative reads derive from the C57BL/6J haplotype, but in spleen the figure is just 39%.

Genes showing divergent allele-specific patterns between tissues were clustered into different functional classes using the DAVID tool⁴¹. Among such genes, those encoding proteins found in mitochondria are significantly enriched between liver and spleen ($\text{FDR} = 9.5 \times 10^{-6}$), as are cell cycle genes between thymus and spleen ($\text{FDR} = 3.4 \times 10^{-4}$), indicating that allele-specific biases are related to the functional program occurring in these tissues.

To characterize the molecular source of allele-specific biases we sequenced DNA from liver bound to chromatin precipitated by a marker for active gene promoters (histone 3, lysine 4 trimethylation; H3K4me3). Of 19,258 SNPs in these ChIP-Seq (chromatin immunoprecipitation followed by sequencing) reads with greater than seven informative reads for H3K4me3, 386 (2%) showed a significant allelic bias. There was unsurprisingly a highly significant correlation between allelic biases of H3K4me3 in the promoters of genes with allelic expression biases ($P < 10^{-10}$). Histone modification of promoter regions, as opposed to other parts of genes, is most predictive of transcriptional bias (Spearman's $\rho = 0.29$), particularly so for the strongly biased genes, showing below 25% or above 75% of the reference allele expression (Spearman's $\rho = 0.67$). Therefore, we have been able to identify genes where differences in *cis*-regulatory promoter sequence between C57BL/6J and DBA/2J are likely to contribute significantly

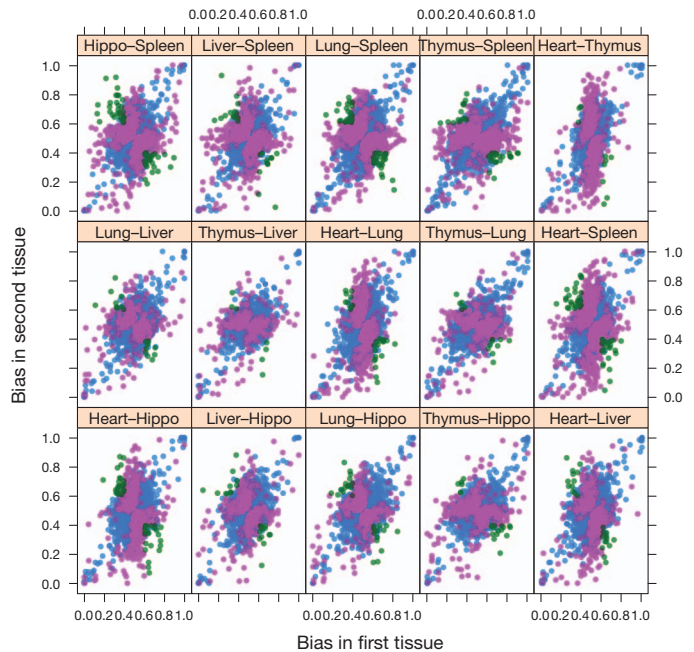


Figure 3 | Allele-specific biases in RNA expression levels between tissues from C57BL/6J x DBA/2J F1 mice. RNA was sequenced from six tissues: hippocampus (hippo), spleen, liver, heart, lung and thymus. Each point represents a gene, and the bias ranges from 1.0 (exclusively C57BL/6J) to 0.0 (exclusively DBA/2J). The tissue comparison is shown above each plot. The points are coloured by whether the difference in bias is not significant (blue), significantly different bias but in the same direction (pink) or significantly different but switching direction (green).

to allele-specific expression biases. With access to the genome sequences, we can use the functionally defined *cis* sequence variants to identify the important regulatory elements.

Molecular basis of quantitative traits

We used the complete genome sequence of multiple inbred strains to address a key challenge in complex trait genetics: the identification of sequence variants that underlie quantitative traits. We asked whether functional variants have common molecular features, and if they were more likely to lie within genes or outside them, and to comprise structural variants, indels or SNPs. We tested the hypothesis that quantitative trait loci (QTLs) with large effects (expressed as the percentage of total phenotypic variation attributable to the locus) are more likely to consist of certain categories of sequence variant.

We examined this relationship using 843 QTLs identified in over 2,000 heterogeneous stock mice, animals that are descended from eight of the sequenced strains (A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP/J)²⁴. Because many recombinants have accumulated in the heterogeneous stock since its creation, the QTLs are resolved to an average genomic size of 3 Mb. The 100 traits mapped include disease models (asthma, anxiety and type 2 diabetes), as well as haematological, immunological, biochemical and anatomical phenotypes^{24,42}.

We imputed the genotypes of the heterogeneous stock mice for all variants and then applied a test that discriminates between variants that could be functional and those that are not⁴³. At each variant we compared two models. In one (the haplotype model) the effect on the QTL was modelled with eight alleles (representing each of the founder haplotypes). In the second, the effect on the QTL was modelled with the number of alleles of the variant (usually two for a SNP). At 718 QTLs (85%) there was at least one variant where the fit of the allelic model was better than a haplotype based model⁴⁴. This implies that, at these QTLs, there is either a single functional variant, or a series of functional variants on the same haplotype. The median number of

Table 2 | The molecular nature of sequence variants and their effect on phenotypic variation.

QTL Pct Var	Intergenic	Downstream	Exon	Intron	Upstream	Coding (detrimental)	SNP	Structural variant	Indel
All	1.18**	0.71	0.7	0.79	0.67	0.79	1.00	0.84	1.04
<4%	1.21**	0.67	0.67	0.75*	0.63	0.74	0.99	0.69**	1.07
>4%	0.57**	1.05	1.28	1.43*	0.97	1.00	1.02	0.85	0.95
>10%	0.65**	1.32	1.59*	1.69**	1.32	2.13*	0.88**	1.69*	1.48**

The class of sequence variants and their position relative to genes influence the likelihood that they are functional, as predicted by a statistical method⁴⁴. The table shows the ratio of variants that score a maximum negative merge log(*P*-value) to those that do not within five different genomic regions: intergenic, exonic, intronic and either 2 kb upstream or downstream of the gene. Ratios are also shown for four molecular types: SNPs, structural variants, insertion/deletions (indels) polymorphisms and SNPs predicted to be detrimental to the coding sequence of a gene. The QTL data used for this analysis were derived from the heterogeneous stock mice²⁴ generated from a cross between eight of the sequenced strains. **P* < 0.05, ***P* < 0.01.

variants per QTL with a merge *P*-value exceeding the minimum haplotype *P*-value was 7; we refer to these variants as functional variants. At 10% of QTLs there is a single functional variant so defined.

We asked whether functional variants are more likely to occur in certain locations relative to genes and whether they are more likely to belong to certain molecular classes. Suppose at a QTL we classify 0.1% of the variants as potentially functional. If there is no relationship between the position of a gene and a functional variant, we expect 0.1% of the variants within genes to be classified as functional. We calculated the ratio of the percentage of functional variants at a QTL over the percentage of variants in five locations relative to genes: intergenic, exonic, intronic or flanking (upstream or downstream lying within 2 kb of the transcriptional start or end sites). Ratios greater than 1 indicate that functional variants are enriched in a classification and less than 1 indicate relative deficiency. We calculated the significance of the ratios' departure from 1 empirically (Table 2). We carried out a similar analysis of molecular categories, comparing SNPs, structural variants, indels and coding polymorphisms predicted to be harmful to protein function.

Figure 4 shows results for 718 QTLs, grouped by effect size (the percentage of phenotypic variance attributed to the QTL) so that each group contains approximately 100 QTLs. We also show results for the 22 largest effect QTLs (explaining more than 10% of the variance). Table 2 shows the results of testing for significant differences between large effect (>4%) and small effect (<4%) QTLs.

Functional variants at small effect QTLs are significantly more likely to be intergenic and less likely to be a structural variant; by contrast, functional variants at large effect QTLs are significantly less likely to be intergenic, and more likely to be intronic. However, it is only with the 3% of QTLs that explain more than 10% of the phenotypic variance that we find significant enrichment for coding variants predicted to be detrimental. These latter QTLs are significantly more likely to arise from indels and structural variants. Our analysis therefore indicates that both the position and molecular nature of quantitative trait variants influence the effect size of the QTL.

Discussion

The sequence we have obtained has a number of notable features. First is the sheer magnitude of the number of variants we have found. An earlier catalogue, based on re-sequencing by hybridization to oligonucleotide arrays, identified SNPs at 8.3M unique sites in 15 strains⁹; our total count in 17 strains is 56.7M unique sites. In addition, our catalogue includes other types of sequence polymorphism that have previously been difficult to assess on a genome-wide scale: indels at 8.8M unique sites and 0.28M structural variants.

Second, we have estimated the false positive and false negative rates by exploiting 17.5 Mb of very high quality sequence from one non-reference strain. We should caution, however, that the BAC sequences were not chosen randomly from the genome; their collinearity when mapped back to the reference genome indicates that they do not lie in regions replete with structural variation for example. Importantly, access to the BAC sequence tells us what the new sequencing technology misses, information currently lacking for other vertebrate sequence projects. We find that inaccessible regions contain almost three times the amount of sequence variation expected from the rate observed in the accessible regions. This observation, gained from the

analysis of inbred genomes that represent a best-case scenario for variant calling, has important implications for the whole genome sequencing of outbred populations such as humans, where variant calling is significantly more difficult.

What use is the current catalogue of variants? First, simply knowing the distribution of variants across the genomes of the sequenced strains is important. The short evolutionary timescale of domestication suggests that many genetic differences among classical inbred strains originated in natural populations. Our phylogenetic analyses both confirm that *M. m. musculus* and *M. m. castaneus* are sister subspecies⁴⁵ and demonstrate that wild mouse genomes are complex mosaics of alternative evolutionary histories. Widespread phylogenetic discordance indicates that polymorphisms are often shared among subspecies, challenging the assignment of subspecific ancestry across the genomes of the classical inbred strains. Our results further suggest that *M. spretus* is not a reliable outgroup for determining the ancestral state of house mice in some genomic regions. Analyses of genome sequences from larger numbers of wild mice will provide a more detailed understanding of the origins of laboratory mice.

A second use of our catalogue is for exploring the relationship between genotype and phenotype. We have demonstrated this with two examples. By examining six tissues in a single cross (C57BL/6J × DBA/2J) we were able to detect high levels of allelic bias at 12% of expressed loci. Furthermore, 4.8% of tested transcripts showed divergent allele-specific patterns between tissues: the allele that is

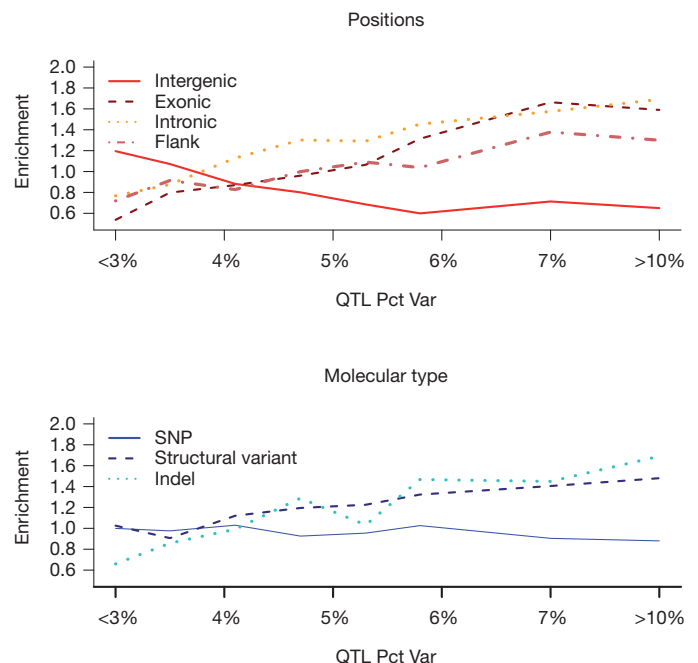


Figure 4 | Enrichment of functional variants. Each line shows the ratio of the percentage of functional variants at a QTL over the percentage of variants expected. Ratios greater than one indicate that functional variants are enriched in a classification and ratios less than one indicate a dearth of functional variants. Functional variants are classified by their position relative to a gene and by their molecular class: SNPs, structural variants and insertion/deletions (indels) polymorphisms.

relatively highly expressed in one tissue is relatively under-expressed in a second tissue. Again using our catalogue, and the genome sequences reported here, we have begun to identify the molecular basis for this complex pattern of gene regulation. Further analysis and functional studies will allow us to identify the exact sequence differences responsible for these allelic expression differences.

We also show that the molecular nature of sequence variants and their position relative to genes influence the likelihood that they are functional. Using a statistical method to predict whether the allelic pattern of a variant is consistent with its action as the molecular cause of quantitative trait variation, we are able to show that functional variants contributing to small effect QTLs are significantly more likely to be intergenic; by contrast, larger effect QTLs are more likely to be caused by intronic variants, and are significantly less likely to be intergenic.

Together with the accumulated phenotypic information on inbred strains, provided by the Mouse Phenome Project, the sequence of the 17 mouse genomes and the associated catalogue of variants will serve as a basis for understanding trait differences, and will allow further insights into the nature of functional variants. Furthermore, near complete sequence will make it possible to impute the genomes of any derivative of the sequenced strains, including the Collaborative Cross²³, a large set of recombinant inbred strains to be used for high-resolution mapping of multiple complex phenotypes. Collectively, the sequence we describe here will help dissect the path from sequence variant to phenotype.

METHODS SUMMARY

The Supplementary Information provides full details of samples, data generation protocols, read mapping, SNP calling, short insertion and deletion calling, structural variation calling and all other computational methods.

Received 5 July; accepted 5 August 2011.

- Paigen, K. One hundred years of mouse genetics: an intellectual history. II. The molecular revolution (1981–2002). *Genetics* **163**, 1227–1235 (2003).
- Paigen, K. One hundred years of mouse genetics: an intellectual history. I. The classical period (1902–1980). *Genetics* **163**, 1–7 (2003).
- Dietrich, W. F. *et al.* Genetic identification of *Mom-1*, a major modifier locus affecting *Min*-induced intestinal neoplasia in the mouse. *Cell* **75**, 631–639 (1993).
- Church, D. M. *et al.* Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol.* **7**, e1000112 (2009).
- Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- van der Weyden, L., Adams, D. J. & Bradley, A. Tools for targeted manipulation of the mouse genome. *Physiol. Genomics* **11**, 133–164 (2002).
- Ringwald, M. *et al.* The IKMC web portal: a central point of entry to data and resources from the International Knockout Mouse Consortium. *Nucleic Acids Res.* **39**, D849–D855 (2011).
- Skarnes, W. C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337–342 (2011).
- Frazer, K. A. *et al.* A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* **448**, 1050–1053 (2007).
- Quinlan, A. R. *et al.* Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. *Genome Res.* **20**, 623–635 (2010).
- Gale, G. D. *et al.* A genome-wide panel of congenic mice reveals widespread epistasis of behavior quantitative trait loci. *Mol. Psychiatry* **14**, 631–645 (2009).
- Iakoubova, O. A. *et al.* Genome-tagged mice (GTM): two sets of genome-wide congenic strains. *Genomics* **74**, 89–104 (2001).
- Bennett, B. J. *et al.* A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Res.* **20**, 281–290 (2010).
- Singer, J. B. *et al.* Genetic dissection of complex traits with chromosome substitution strains of mice. *Science* **304**, 445–448 (2004).
- Shao, H. *et al.* Analyzing complex traits with congenic strains. *Mamm. Genome* **21**, 276–286 (2010).
- Hunter, K. W. & Crawford, N. P. The future of mouse QTL mapping to diagnose disease in mice in the age of whole-genome association studies. *Annu. Rev. Genet.* **42**, 131–141 (2008).
- Rozmahe, R. *et al.* Modulation of disease severity in cystic fibrosis transmembrane conductance regulator deficient mice by a secondary genetic factor. *Nature Genet.* **12**, 280–287 (1996).
- Shao, H. *et al.* Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. *Proc. Natl Acad. Sci. USA* **105**, 19910–19914 (2008).
- Flint, J. & Mackay, T. F. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* **19**, 723–733 (2009).
- Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
- Turner, D. J., Keane, T. M., Sudbery, I. & Adams, D. J. Next-generation sequencing of vertebrate experimental organisms. *Mamm. Genome* **20**, 327–338 (2009).
- Guan, C., Ye, C., Yang, X. & Gao, J. A review of current large-scale mouse knockout efforts. *Genesis* **48**, 73–85 (2010).
- Churchill, G. A. *et al.* The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nature Genet.* **36**, 1133–1137 (2004).
- Valdar, W. *et al.* Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nature Genet.* **38**, 879–887 (2006).
- Davis, R. C. *et al.* A genome-wide set of congenic mouse strains derived from DBA/2J on a C57BL/6J background. *Genomics* **86**, 259–270 (2005).
- Steward, C. A. *et al.* Genome-wide end-sequenced BAC resources for the NOD/MarkTac and NOD/ShiLtJ mouse genomes. *Genomics* **95**, 105–110 (2010).
- Sherry, S. T., Ward, M. & Sirotkin, K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.* **9**, 677–679 (1999).
- Yalcin, B. *et al.* Sequence-based characterization of structural variation in the mouse genome. *Nature* doi:10.1038/nature10432 (this issue).
- Deem, A. *et al.* Break-induced replication is highly inaccurate. *PLoS Biol.* **9**, e1000594 (2011).
- Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C. & Forejt, J. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323**, 373–375 (2009).
- McDonald, J. P. *et al.* 129-derived strains of mice are deficient in DNA polymerase ϵ and have normal immunoglobulin hypermutation. *J. Exp. Med.* **198**, 635–643 (2003).
- Koike, H., Arguello, P. A., Kvajo, M., Karayiorgou, M. & Gogos, J. A. Disc1 is mutated in the 129SvEv strain and modulates working memory in mice. *Proc. Natl Acad. Sci. USA* **103**, 3693–3697 (2006).
- Lutzmann, M. & Mechali, M. How to load a replicative helicase onto chromatin: a more and more complex matter during evolution. *Cell Cycle* **8**, 1309–1313 (2009).
- Mural, R. J. *et al.* A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science* **296**, 1661–1671 (2002).
- Ané, C., Larget, B., Baum, D. A., Smith, S. D. & Rokas, A. Bayesian estimation of concordance among gene trees. *Mol. Biol. Evol.* **24**, 412–426 (2007).
- Pamilo, P., Nei, M. & Li, W. H. Accumulation of mutations in sexual and asexual populations. *Genet. Res.* **49**, 135–146 (1987).
- Rosenberg, N. A. The probability of topological concordance of gene trees and species trees. *Theor. Popul. Biol.* **61**, 225–247 (2002).
- Baum, D. A. Concordance trees, concordance factors, and the exploration of reticulate genealogy. *Taxon* **56**, 417–426 (2007).
- Laurie, C. C. *et al.* Linkage disequilibrium in wild mice. *PLoS Genet.* **3**, e144 (2007).
- McDaniell, R. *et al.* Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **328**, 235–239 (2010).
- Huang, D. W. *et al.* Extracting biological meaning from large gene lists with DAVID. *Curr. Protoc. Bioinformatics* **13**, Unit 13.11 (2009).
- Solberg, L. C. *et al.* A protocol for high-throughput phenotyping, suitable for quantitative trait analysis in mice. *Mamm. Genome* **17**, 129–146 (2006).
- Yalcin, B., Flint, J. & Mott, R. Using progenitor strain information to identify quantitative trait nucleotides in outbred mice. *Genetics* **171**, 673–681 (2005).
- Mott, R., Talbot, C. J., Turri, M. G., Collins, A. C. & Flint, J. A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc. Natl Acad. Sci. USA* **97**, 12649–12654 (2000).
- White, M. A., Ané, C., Dewey, C. N., Larget, B. R. & Payseur, B. A. Fine-scale phylogenetic discordance across the house mouse genome. *PLoS Genet.* **5**, e1000729 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This project was supported by The Medical Research Council, UK, and the Wellcome Trust. D.J.A. is supported by Cancer Research UK and L.R., L.R.-D. and A.C. were supported by the Jackson Laboratory. B.A.P. was supported by NSF Grant DEB 0918000 and M.A.W. was supported by an NLM training grant in Computational and Informatics in Biology and Medicine to the University of Wisconsin (NLM 2T15LM007359). R.M., L.G. were supported by Wellcome Trust Grants 085906/Z/08/Z and 083573/Z/07/Z and BBSRC grant BB/F022697/1. The NOD/ShiLtJ BAC sequencing and the next generation Illumina sequencing was funded by Immune Tolerance Network Contract AI 15416, which was sponsored by the National Institute of Allergy and Infectious Diseases, the National Institute of Diabetes and Digestive and Kidney Diseases, and Juvenile Diabetes Research Foundation International. We thank staff in the Sanger Institute sequencing and informatics teams for making this project possible.

Author Contributions D.J.A. and J.F. conceived the study, directed the research, and wrote the paper. T.M.K., P.D., L.G., B.P., M.W., K.W., B.Y., A.H., A.A., G.S., M.G., N.F., E.E., C.N., H.W., J.C., D.J., P.H.-P., A.B., J.N., X.G., W.Y., A.B., L.v.d.W., C.A.S., S.B., J.S., R.M., R.D., I.J., C.P.P. and E.B. performed data analysis. L.R., A.C. and L.D. provided essential biological resources.

Author Information Genomic structural variant study data is deposited in dbSNP (Handle: SC_MOUSE_GENOMES) and DGVA (estd118). Sequence accession numbers are provided in the Supplementary Information. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.F. (jf@well.ox.ac.uk) or D.J.A. (da1@sanger.ac.uk).

The Sagittarius impact as an architect of spirality and outer rings in the Milky Way

Chris W. Purcell^{1,2}, James S. Bullock², Erik J. Tollerud², Miguel Rocha² & Sukanya Chakrabarti³

Like many galaxies of its size, the Milky Way is a disk with prominent spiral arms rooted in a central bar¹, although our knowledge of its structure and origin is incomplete. Traditional attempts to understand our Galaxy's morphology assume that it has been unperturbed by major external forces. Here we report simulations of the response of the Milky Way to the infall of the Sagittarius² dwarf galaxy (Sgr), which results in the formation of spiral arms, influences the central bar and produces a flared outer disk. Two ring-like wrappings emerge towards the Galactic anti-Centre in our model that are reminiscent of the low-latitude arcs observed in the same area of the Milky Way. Previous models have focused on Sgr itself^{3,4} to reproduce the dwarf's orbital history and place associated constraints on the shape of the Milky Way gravitational potential, treating the Sgr impact event as a trivial influence on the Galactic disk. Our results show that the Milky Way's morphology is not purely secular in origin and that low-mass minor mergers predicted to be common throughout the Universe⁵ probably have a similarly important role in shaping galactic structure.

To discern the specific effect of the Sgr impact on the Galactic disk, we need to simulate directly the dark matter and stellar components in both the Milky Way and the Sgr progenitor and to ensure that Sgr has a realistic dark-to-baryonic mass ratio, given the Λ CDM (where Λ represents the accelerating expansion of our Universe, which has a matter density dominated by Cold Dark Matter) cosmology's prediction that even small dwarf galaxies are hosted by massive halos of dark matter. Given the total luminosity (a few times $10^8 L_\odot$, where L_\odot is the solar luminosity) of the Sgr tidal stream and remnant core⁶, cosmological abundance matching demands that the original mass of the dwarf galaxy progenitor was at least $10^{10.5} M_\odot$ (where M_\odot is the solar mass), although best estimates^{7,8} place it in a much more massive halo of roughly $10^{11} M_\odot$. A recent dynamical analysis finds comparable masses, noting that the future discovery of additional stellar debris in the Sgr stream would tend to support the heavier value⁶. We therefore adopt the two masses mentioned above as lower and upper limits, which we refer to as our 'light Sgr' and 'heavy Sgr' models. Our initial Milky Way disk model⁹ matches theoretical expectations and the observed characteristics of the Galaxy.

In isolation, our modelled primary disk begins to form a weak bar after about two billion years, but otherwise remains quite smooth beyond about 5 kiloparsecs (kpc) from its centre. In contrast, the Sgr interactions provide significant perturbations to the outer disk, triggering the formation of outer rings of stellar material and influencing the evolution and formation of the central bar and inner spirality. Each of the model Sgr progenitors experiences two disk crossings, approaching a third at the present day, and the response of the disk is similar in both cases as shown in Fig. 1. The satellite first crosses the disk at a Galactic Centre distance of about 30 kpc approximately 1.75 billion years ago, producing the most significant perturbation. The progenitor loses roughly 75% of its dark matter mass (but little stellar material) during this time, and the disk experiences a caustic signature initially pointing towards the encounter but eventually shearing into trailing spiral ring-like structure; see Fig. 2. The second crossing incites a

weaker ancillary arm with a pitch angle different from that of the primary mode and begins to liberate stellar material from Sgr. These repeated polar encounters produce flaring, asymmetric sloshing in the disk plane, and vertical oscillations above and below the plane of the forming spiral wraps.

The evolution of the central bar can also be affected by perturbing impacts. Although bar formation is sensitive to initial conditions, it is interesting to compare results from run to run, which rely on identical

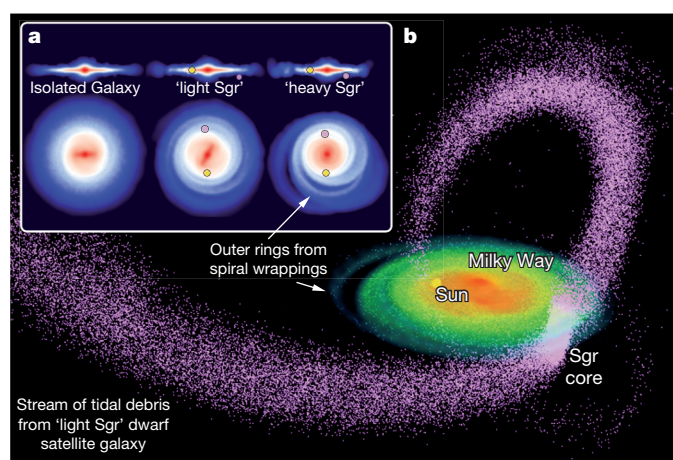


Figure 1 | Visualizations of evolved disk end states in the simulation suite. **a**, Edge- and face-on surface density depictions for each infall model as well as an isolated Galaxy model subject only to secular evolution. The Sun's location is marked as a yellow circle and the present location of the Sgr remnant is marked as a pink circle. The primary Milky Way analogue was initialized via self-consistent multi-component distribution functions⁹ and proved fairly robust to secular instabilities, as shown in the left image after about 2.7 billion years of isolated evolution. **b**, Global rendering of the 'light Sgr' tidal debris and the Milky Way disk. The primary galaxy included a Navarro-Frenk-White (NFW) dark halo¹⁹ with scale radius $r_s = 14.4$ kpc and virial mass $M_{\text{vir}} = 10^{12} M_\odot$; the disk had a mass of $3.59 \times 10^{10} M_\odot$, an exponential scale length of 2.84 kpc and a vertical sech² scale height of 0.43 kpc; the central bulge had a mass of $9.52 \times 10^9 M_\odot$ and an $n = 1.28$ Sérsic profile with a 0.56-kpc effective radius. The 'light Sgr' (or 'heavy Sgr') progenitor with effective virial mass $M_{\text{vir}} = 10^{10.5} M_\odot$ (or $10^{11} M_\odot$) was initialized with an NFW dark halo of scale length 4.9 kpc (or 6.5 kpc) self-consistently with a separate stellar component²⁰ motivated by an analysis of the observed Sgr debris and core⁶: a King profile²¹ with core radius 1.5 kpc, tidal radius 4.0 kpc, and central velocity dispersion equal to 23 km s^{-1} (or 30 km s^{-1}). Following previous work on the Sgr interaction²², our satellites started 80 kpc from the Galactic Centre in the plane of the Milky Way, travelling vertically at 80 km s^{-1} towards the north Galactic pole. We account for the mass loss that would have occurred between virial-radius infall and this 'initial' location by truncating the Sgr progenitor NFW mass profile at the instantaneous Jacobi tidal radius, $r_t = 23.2$ kpc for 'light Sgr' (or 30.6 kpc for 'heavy Sgr'), leaving a total bound mass that is a factor of approximately three smaller than their effective virial mass derived from abundance matching. All simulations used the parallel N -body tree code ChaNGa with a gravitational softening length of one parsec, and followed the evolution of 30 million particles with masses in the range $1.1\text{--}1.9 \times 10^4 M_\odot$.

¹Department of Physics and Astronomy, The University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. ²Center for Cosmology, Department of Physics and Astronomy, The University of California, Irvine, California 92697, USA. ³Department of Physics, Florida Atlantic University, Boca Raton, Florida 33431, USA.

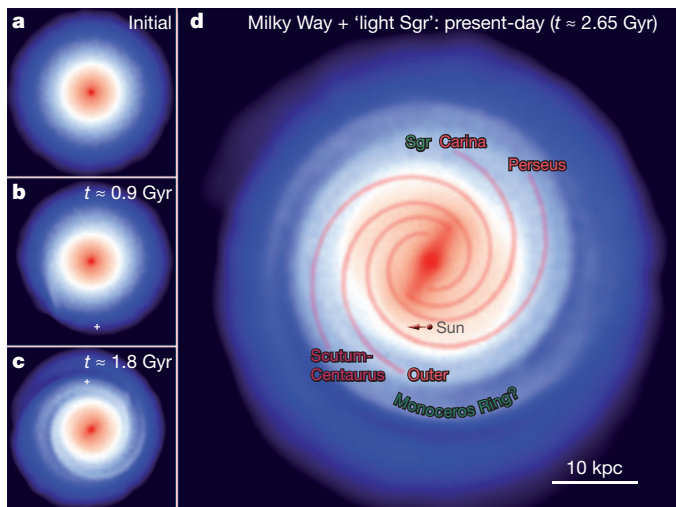


Figure 2 | Face-on surface density visualizations of the Milky Way at four important moments during the ‘light Sgr’ simulation. **a**, Initial model. **b**, Immediately after the first pericentre, where the white cross marks the Sgr impact point. **c**, Shortly after the second pericentric disk crossing. **d**, At the present-day (corresponding to elapsed simulation time 2.65 billion years), overlaid by a four-armed symmetric-spiral fit to the observed arms of the Milky Way as revealed by mapping neutral hydrogen²³. The traditional view of the Milky Way as a secularly evolving system has encouraged theoretical descriptions of quasi-stationary density-wave spirality, although the large peculiar motions of young stars in spiral arms support a more transient picture²⁴ (numerical evidence exists for both short-lived configurations²⁵ as well as more stable forms of spirality, varying with the strength of the tidal induction²⁶). Dynamical analysis of each impacted Milky Way model reveals the importance of the swing amplification mechanism, in which gravitational disturbances in the stellar disk at each pericentric approach shear into trailing arms that are subsequently enhanced on small scales (even in a globally stable system), strengthening transient spiral modes. (Gyr, billion years.)

primary disk models. Compared to our isolated run, the ‘light Sgr’ model induces a more pronounced bar with a faster angular speed. Our ‘heavy Sgr’ case suppresses bar formation compared to the isolated run, as a result of enhanced central disk heating. Although the bar grows with time in the isolated case, at fixed time the ‘light Sgr’ run always produces a more pronounced bar and the ‘heavy Sgr’ run always produces a less pronounced bar. Both Sgr-infall models each have an end-state bar orientation ($\phi_{\text{bar}} \approx 15\text{--}20^\circ$) that corresponds to estimates¹⁰ of the long bar at the centre of the Milky Way ($\phi_{\text{MW}} \approx 15\text{--}30^\circ$). Our isolated run does not, being phase-shifted from the impacted bars by roughly 90° , which indicates that the Sgr event must be considered in any model that attempts to detail the evolution of the Galactic bar.

A vital test of the model’s viability is the preservation of a disk that is as thin and dynamically cold as the Milky Way. Though our resultant disks do show flaring at large radius, the scale heights remain less than 0.5 kpc well beyond the solar radius for both model cases. The velocity ellipsoids of the remnant disks in the solar vicinity are $(\sigma_R, \sigma_\phi, \sigma_z) \approx (37, 27, 20) \text{ km s}^{-1}$ for the ‘light Sgr’ case and $(33, 42, 22) \text{ km s}^{-1}$ for the ‘heavy Sgr’ case, which are grossly consistent with constraints¹¹ placed on nearby stars with ages of about 4–8 billion years, that is, $(\sigma_R, \sigma_\phi, \sigma_z) \approx (35, 25, 20) \text{ km s}^{-1}$.

At the present time in the real Galaxy (and in the infall models, as shown in Fig. 3 for the ‘light Sgr’ case), the Sgr core is moving up towards the Galactic plane¹² and has two distinct tidal arms resulting from its advanced stage of disruption. The orbital plane of Sgr allows us to fix the Galactic longitude of the Sgr remnant in our simulations at $l \approx 5.6^\circ$, and we establish the endstate time step when the dwarf core is at a Galactic latitude of $b \approx -14.0^\circ$. These coordinates are a good match to observations^{2,13}; the heliocentric distance of our satellite remnant is around 22 kpc for the ‘light Sgr’ model and around 20 kpc for the ‘heavy Sgr’ model, commensurate with the $24 \pm 4 \text{ kpc}$ range typically

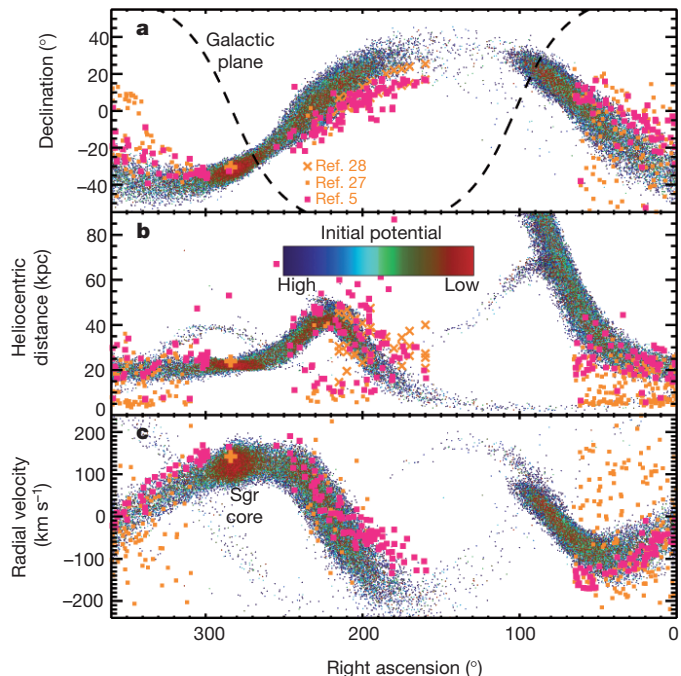


Figure 3 | The observed Sgr tidal debris stream and remnant core in comparison to our ‘light Sgr’ simulation, in equatorial coordinates. **a**, Declination versus right ascension. **b**, Heliocentric distance versus right ascension. **c**, Radial velocity versus right ascension. Simulated particles are coloured according to their initial potential energy, and the orange points are data from 2MASS M-giant stars²⁷ and Sloan Digital Sky Survey red-clump stars²⁸ (marked by squares and crosses respectively; thick crosses denote canonical values for the remnant core^{2,12}). The pink points are 2MASS M-giants identified as likely stream members⁴. The present-day location of the simulated remnant and tidal arms are similar to those observed. Combining this with observational constraints on the dispersion ($\sigma \approx 10\text{--}15 \text{ km s}^{-1}$), breadth (8–10 kpc; ref. 27), and length of the observed debris stream provides some legitimacy for our model.

derived¹⁴. The stellar velocity dispersion of both the core and the stream in our remnants are consistent with measurements for Sgr^{4,15} (approximately $10\text{--}20 \text{ km s}^{-1}$), although precise results are sensitive to stellar initial conditions. Our simulated Sgr debris distributions do not precisely match all of the observed characteristics, but we argue that these differences are not significant enough to alter our gross expectation that the Sgr impact has significantly affected the Milky Way disk, given that dark matter in the progenitor is the main driver of disk perturbations. Better constraints on debris stream dispersion, length, and thickness may provide a way of constraining the full progenitor mass in the future.

The disks in our simulations develop outer arcs of material generated in association with each disk crossing. These evolved outer wrappings are loosely wound and resemble rings. One of the predicted arcs, at about 10 kpc from the Sun, is reminiscent of the low-latitude Milky Way feature known as the Monoceros ring. Though the Monoceros ring is often considered to be the leftover tidal stream from a now-defunct dwarf satellite galaxy¹⁶, some observational evidence has suggested that the Monoceros ring could be a feature of the Milky Way itself⁷. Previous theoretical work has suggested that a past encounter with some previously unidentified massive satellite could have produced the Monoceros ring as the outcome of a disk impact¹⁸. We specifically identify the Sgr progenitor as the likely candidate for the impact that moulded the Monoceros ring from the Milky Way disk, as the induced spiral arms detached from the outer Galactic plane and began to oscillate vertically over a range of 5–10 kpc (see Fig. 4).

Finally, we note that these predicted ring-like wrappings of already-known spiral arms will also be potentially observable at deep magnitudes by next-generation mapping surveys. These efforts may connect the features in the Galactic outskirts to the global structure of the Milky Way disk, further implicating the Sgr dwarf as a principal shaper of

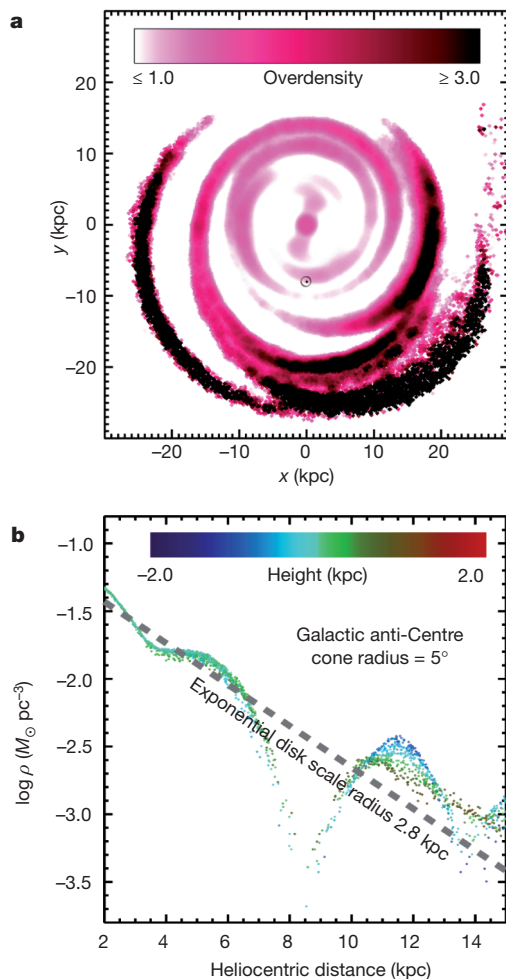


Figure 4 | Endstate disk overdensities in the ‘heavy Sgr’ simulation. **a**, Disk stellar overdensity, colour-coded by the ratio of three-dimensional stellar density to the local axisymmetric mean density of the disk. **b**, Local stellar density in a thin cone directed from the solar neighbourhood towards the - Galactic anti-Centre, as a function of heliocentric distance along the cone. In both panels, off-plane overdensities resemble ‘multiple tributaries’ observed in the Milky Way²⁹, and the spiral arm wrapping at a distance of around 10 kpc is strikingly similar to the Monoceros ring. The Monoceros ring feature spans a wide range in metallicity³⁰: $[\text{Fe}/\text{H}] \approx -1.6$ to -0.4 . The corresponding high-latitude arc in our simulation at that distance is composed of stars that were initially widely distributed throughout the disk, suggesting that radial mixing during the Sgr impact must be an important factor in the Milky Way’s recent chemodynamical evolution, and also that the chemical composition of the real Monoceros ring feature may not be as reliable in discriminating models for the origin of the ring as otherwise expected. See the Supplementary Information, in which we discuss the quantitative agreement between our Galactic anti-Centre spiral structure and Monoceros ring observations in more detail.

Milky Way morphology. Cosmological considerations strongly suggest that the Sgr progenitor was massive, and thus motivate the expectation that it has influenced Galactic evolution. More broadly, the implication that Sgr has affected the Milky Way morphology provides an indication that minor mergers shape galactic structure throughout the Universe. Future observations of the kinematics and extent of the Sgr tidal debris stream will further constrain the scenario discussed here and shape the perspective that the Sgr impact must be included in future theories of Milky Way evolution.

Received 1 April; accepted 1 August 2011.

- Pohl, M., Englmaier, P. & Bissantz, N. Three-dimensional distribution of molecular gas in the barred Milky Way. *Astrophys. J.* **677**, 283–291 (2008).
- Ibata, R. A., Gilmore, G. & Irwin, M. J. A dwarf satellite galaxy in Sagittarius. *Nature* **370**, 194–196 (1994).

- Law, D. R., Johnston, K. V. & Majewski, S. R. A Two Micron All-Sky Survey view of the Sagittarius dwarf galaxy. IV. Modeling the Sagittarius tidal tails. *Astrophys. J.* **619**, 807–823 (2005).
- Law, D. R. & Majewski, S. R. The Sagittarius dwarf galaxy: a model for evolution in a triaxial Milky Way halo. *Astrophys. J.* **714**, 229–254 (2010).
- Johnston, K. V., Spergel, D. N. & Hernquist, L. The disruption of the Sagittarius dwarf galaxy. *Astrophys. J.* **451**, 598–606 (1995).
- Niederste-Ostholt, M., Belokurov, V., Evans, N. W. & Peñarrubia, J. Re-assembling the Sagittarius dwarf galaxy. *Astrophys. J.* **712**, 516–526 (2010).
- Conroy, C. & Wechsler, R. H. Connecting galaxies, halos, and star formation rates across cosmic time. *Astrophys. J.* **696**, 620–635 (2009).
- Behroozi, P. S., Conroy, C. & Wechsler, R. H. A comprehensive analysis of uncertainties affecting the stellar mass-halo mass relation for $0 < z < 4$. *Astrophys. J.* **717**, 379–403 (2010).
- Widrow, L. M., Pym, B. & Dubinski, J. Dynamical blueprints for galaxies. *Astrophys. J.* **679**, 1239–1259 (2008).
- Bissantz, N. & Gerhard, O. Spiral arms, bar shape and bulge microlensing in the Milky Way. *Mon. Not. R. Astron. Soc.* **330**, 591–608 (2002).
- Nordström, B. *et al.* The Geneva-Copenhagen survey of the solar neighbourhood. Ages, metallicities, and kinematic properties of $\sim 14,000$ F and G dwarfs. *Astron. Astrophys.* **418**, 989–1029 (2004).
- Dinescu, D. I., Girard, T. M., van Altena, W. F. & López, C. E. Absolute Proper Motion of the Sagittarius Dwarf Galaxy and of the Outer Regions of the Milky Way Bulge. *Astrophys. J.* **618**, L25–L28 (2005).
- Majewski, S. R., Skrutskie, M. F., Weinberg, M. D. & Ostheimer, J. C. A Two Micron All Sky Survey view of the Sagittarius dwarf galaxy. I. Morphology of the Sagittarius core and tidal arms. *Astrophys. J.* **599**, 1082–1115 (2003).
- Kunder, A. & Chaboyer, B. Distance to the Sagittarius dwarf galaxy using Macho Project RR Lyrae stars. *Astron. J.* **137**, 4478–4486 (2009).
- Lokas, E. L. *et al.* The inner structure and kinematics of the Sagittarius dwarf galaxy as a product of tidal stirring. *Astrophys. J.* **725**, 1516–1527 (2010).
- Peñarrubia, J. *et al.* A comprehensive model for the Monoceros tidal stream. *Astrophys. J.* **626**, 128–144 (2005).
- Momany, Y. *et al.* Outer structure of the Galactic warp and flare: explaining the Canis Major over-density. *Astron. Astrophys.* **451**, 515–538 (2006).
- Kazantzidis, S., Bullock, J. S., Zentner, A. R., Kravtsov, A. V. & Moustakas, L. A. Cold Dark Matter substructure and galactic disks. I. Morphological signatures of hierarchical satellite accretion. *Astrophys. J.* **688**, 254–276 (2008).
- Navarro, J. F., Frenk, C. S. & White, S. D. M. The structure of cold dark matter halos. *Astrophys. J.* **462**, 563–575 (1996).
- Zemp, M., Moore, B., Stadel, J., Carollo, C. M. & Madau, P. Multimass spherical structure models for N-body simulations. *Mon. Not. R. Astron. Soc.* **386**, 1543–1556 (2008).
- King, I. R. The profile of a star image. *Publ. Astron. Soc. Pac.* **83**, 199–201 (1971).
- Keselman, J. A., Nusser, A. & Peebles, P. J. E. Galaxy satellites and the weak equivalence principle. *Phys. Rev. D* **80**, 063517 (2009).
- Levine, E. S., Blitz, L. & Heiles, C. The spiral structure of the outer Milky Way in hydrogen. *Science* **312**, 1773–1777 (2006).
- Reid, M. J. *et al.* Trigonometric parallaxes of massive star-forming regions. VI. Galactic structure, fundamental parameters, and noncircular motions. *Astrophys. J.* **700**, 137–148 (2009).
- Sellwood, J. A. The lifetimes of spiral patterns in disc galaxies. *Mon. Not. R. Astron. Soc.* **410**, 1637–1646 (2010).
- Struck, C., Dobbs, C. L. & Hwang, J.-S. Slowly-breaking waves: the longevity of tidally induced spiral structure. *Mon. Not. R. Astron. Soc.* **414**, 2498–2510 (2011).
- Majewski, S. R. *et al.* A Two Micron All Sky Survey view of the sagittarius dwarf galaxy. II. Swope telescope spectroscopy of M giant stars in the dynamically cold Sagittarius tidal stream. *Astron. J.* **128**, 245–259 (2004).
- Correnti, M., Bellazzini, M., Ibata, R. A., Ferraro, F. R. & Varghese, A. The northern wraps of the Sagittarius stream as traced by red clump stars: distances, intrinsic widths, and stellar densities. *Astrophys. J.* **721**, 329–356 (2010).
- Grillmair, C. J. Substructure in tidal streams: tributaries in the anticenter stream. *Astrophys. J.* **651**, L29–L32 (2006).
- Ibata, R. A., Irwin, M. J., Lewis, G. F., Ferguson, A. M. N. & Tanvir, N. One ring to encompass them all: a giant stellar structure that surrounds the Galaxy. *Mon. Not. R. Astron. Soc.* **340**, L21–L27 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank K. Johnston, M. Kaplinghat, D. Law, H. Morrison and A. Zentner for discussions and C. Struck for suggestions that improved this work. All simulations were performed on the GreenPlanet cluster at the University of California, Irvine.

Author Contributions C.W.P. helped to conceive the project, performed and analysed all simulations, and wrote the majority of the text. J.S.B. helped to conceive the project, contributed to the analysis, and co-authored the text. E.J.T. created the 3D visualizations and provided discussion and direction related to observational correlations. M.R. provided the code that was utilized to initialize the Sagittarius progenitors and insight on how to thereby achieve the desired properties of these systems. S.C. assisted with interpretation and analysis of the stellar disk instabilities and their time evolution.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.W.P. (cpurcell@pitt.edu).

Geometrical enhancement of low-field magnetoresistance in silicon

Caihua Wan^{1,2}, Xiaozhong Zhang^{1,2}, Xili Gao^{1,2}, Jimin Wang^{1,2} & Xinyu Tan^{1,2}

Inhomogeneity-induced magnetoresistance (IMR) reported in some non-magnetic semiconductors^{1–8}, particularly silicon^{1,6–8}, has generated considerable interest owing to the large magnitude of the effect and its linear field dependence (albeit at high magnetic fields). Various theories implicate^{9–18} spatial variation of the carrier mobility as being responsible for IMR. Here we show that IMR in lightly doped silicon can be significantly enhanced through hole injection, and then tuned by an applied current to arise at low magnetic fields. In our devices, the ‘inhomogeneity’ is provided by the p–n boundary formed between regions where conduction is dominated by the minority and majority charge carriers (holes and electrons) respectively; application of a magnetic field distorts the current in the boundary region, resulting in large magnetoresistance. Because this is an intrinsically spatial effect, the geometry of the device can be used to enhance IMR further: we designed an IMR device whose room-temperature field sensitivity at low fields was greatly improved, with magnetoresistance reaching 10% at 0.07 T and 100% at 0.2 T, approaching the performance of commercial giant-magnetoresistance devices^{19,20}. The combination of high sensitivity to low magnetic fields and large high-field response should make this device concept attractive to the magnetic-field sensing industry. Moreover, because our device is based on a conventional silicon platform, it should be possible to integrate it with existing silicon devices and so aid the development of silicon-based magnetoelectronics.

We measured the current–voltage (*I*–*V*) characteristics of an Si wafer with intrinsic SiO₂ on its top surface (sample 15) using the four-electrode method (Fig. 1a). The SiO₂ layer aided hole injection (see Supplementary Information). The *I*–*V* curve (Fig. 1b) was separated into three regions according to their resistivities. Regions 1 and 2 have resistivities of 31 Ω m and 153 Ω m, respectively, and between these lies a transition region. The Hall coefficient of sample 15 (Fig. 1b) changed from $-3.79 \text{ m}^3 \text{ C}^{-1}$ to $1.55 \text{ m}^3 \text{ C}^{-1}$ over the transition region. The carrier type in the sample was inverted from majority (electrons) to minority (holes) with increasing applied current. The carrier densities *n* and *p* were about $-1.7 \times 10^{12} \text{ cm}^{-3}$ and $4.0 \times 10^{12} \text{ cm}^{-3}$ for regions 1 and 2, respectively.

The holes, once injected, drift into the inside of the sample under electric field until their recombination with electrons. The distribution of holes can be modelled as $p = p_0 \exp(-r/\mu_h E \tau_h)$ where p_0 is the hole density near the injection electrode, *p* is its counterpart at a distance *r* from the injecting electrode, *E* is electric field, μ_h is mobility and τ_h is lifetime (see Supplementary Information). The holes clustered mainly in a region a distance $r < r_0 \equiv \mu_h E \tau_h$ from the electrode, where r_0 characterizes the length of the minority region. Thus, the p–n boundary formed between the majority and minority regions could be moved forward along the *x* axis, because $r_0 \propto E$. r_0 was about 2.0–4.0 mm at the transition region, similar to *D*, indicating that hole injection and carrier type inversion can be detected using *I*–*V* and Hall measurements. The insets in Fig. 1 schematically show the location of the p–n boundary moving from region 1, through the transition region to region 2 with increasing applied current.

We found that the *I*–*V* characteristics of sample 20 were strongly affected by magnetic field (Fig. 2a). The transition region of 146–226 μA at 0 T was shifted downward to 80–130 μA at 7 T, accompanied by enhanced magnetoresistance (Fig. 2b). Here we define the magnetoresistance as $[R(B) - R(B = 0)]/R(B = 0) = [V(B) - V(B = 0)]/V(B = 0)$. The magnetoresistance of sample 20 exhibited three kinds of field dependence in the three regions. In region 1 the magnetoresistance reached 56.6% at 7 T and is equal to $(\mu_e B)^2$, where μ_e is electron mobility. This normal magnetoresistance originated from the curvature of the carrier trajectory under magnetic field. The fit gives $\mu_e \approx 0.11 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$, close to our calibrated value of $0.12 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$. In the transition region of 146–226 μA, the magnetoresistance changed from normal magnetoresistance to an ‘abnormal’ magnetoresistance marked by giant magnitude and a field

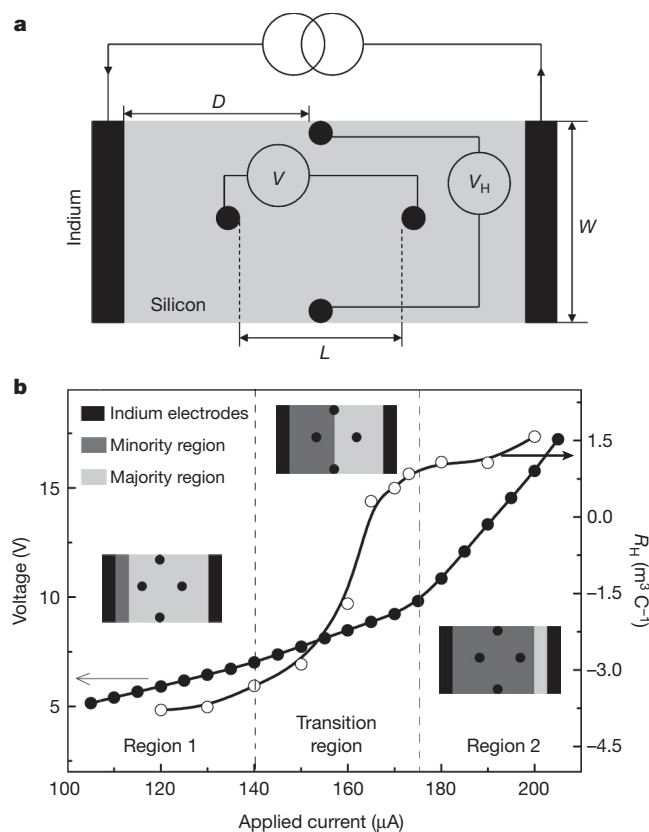


Figure 1 | *I*–*V* characteristics and Hall coefficient measured in In/SiO₂/Si/SiO₂/In at 300 K. a, Measurement geometry. The width *W* is 3.0 mm, the distance between the voltage electrodes *L* is 2.3 mm and the lateral distance between the current injecting electrode and the Hall electrodes *D* is 3.2 mm. **b**, The closed and open circles show the *I*–*V* characteristics and Hall coefficient of sample 15, respectively. Insets show the locations of the p–n boundary.

¹Laboratory of Advanced Materials, Department of Materials Science and Engineering, Tsinghua University, Beijing 100084, China. ²Beijing National Center for Electron Microscopy, Tsinghua University, Beijing 100084, China.

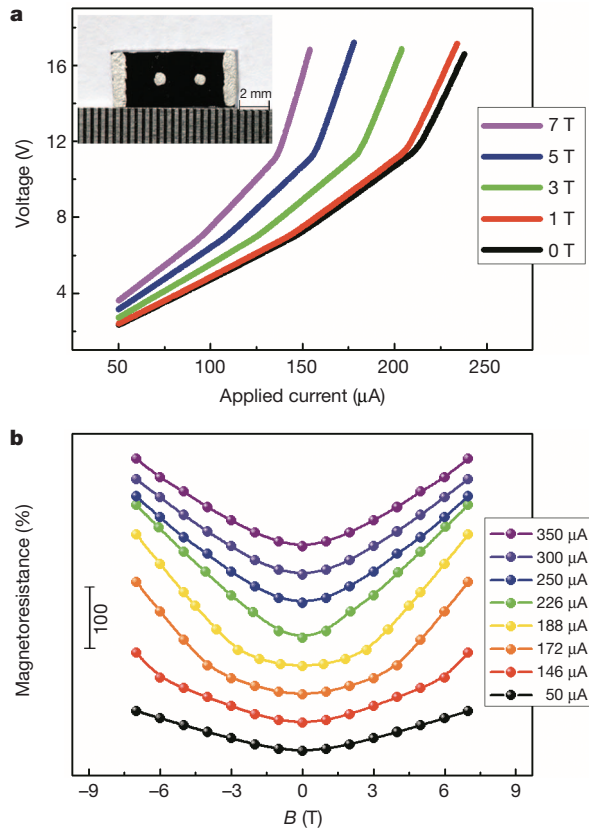


Figure 2 | *I*–*V* characteristics under magnetic field and the magnetoresistance of sample 20 at 300 K. **a, *I*–*V* characteristics of sample 20 under different magnetic fields. The current range of the transition region shifted towards a lower-current region as the magnetic field increased. The inset shows the geometry of the sample. **b**, The magnetic field dependence of magnetoresistance in sample 20. Between 146 μA and 226 μA, a magnetoresistance transition occurred from normal to abnormal. Above the transition, abnormal magnetoresistance was also observed.**

dependence differing from $(\mu B)^2$. The largest magnetoresistance we observed was about 190% at 226 μA and 7 T. Although the crossover phenomenon from parabolic to linear field dependence has also been observed in other systems^{1,7}, there are few ways to control the transition. The crossover field B_0 in our case, however, could be controlled by applied current in the transition region, decreasing from 4.2 T at 146 μA to 0.9 T at 226 μA. In region 2, as current was increased further, we observed only abnormal magnetoresistance. The magnetoresistance at 7 T decreased from 160% at 250 μA to 120% at 350 μA. The largest abnormal magnetoresistance appeared in the transition region, indicating its close relation with the carrier-type inversion induced by hole injection. The IMR theories^{9–11} predicted that in co-existing hole and electron systems IMR would be enhanced by a large spatial variance of mobility $\Delta\mu$ once $\Delta\mu > \mu_{av}$ with $B_0 \propto [\max(\mu_{av}, \Delta\mu)]^{-1}$. In our case, the average mobility $\mu_{av} = (\mu_e + \alpha\mu_h)/(1 + \alpha)$ where $\alpha \equiv p/n$. Therefore $\Delta\mu = \alpha^{1/2}/(1 + \alpha)[\mu_e - \mu_h]$, where p , μ_h and n are hole density, mobility and electron density, respectively. Hole injection at the transition region increased α until $\alpha = 1$, leading to a significant increase in $\Delta\mu$. As a result, B_0 decreased sharply and the magnetoresistance magnitude increased accordingly (see Supplementary Information).

To explore the origin of the IMR further, we analysed the potential and current density distribution of samples using finite element modelling. Hole injection was taken as a prerequisite in our modelling. For simplicity, we let $n = p$, $\mu_e = \mu_h = \mu$ and placed the p–n boundary on the y axis (Fig. 3a). As soon as current was injected from the left electrode, potential built up and distributed symmetrically about the x axis. Once a magnetic field was applied normal to the sample plane,

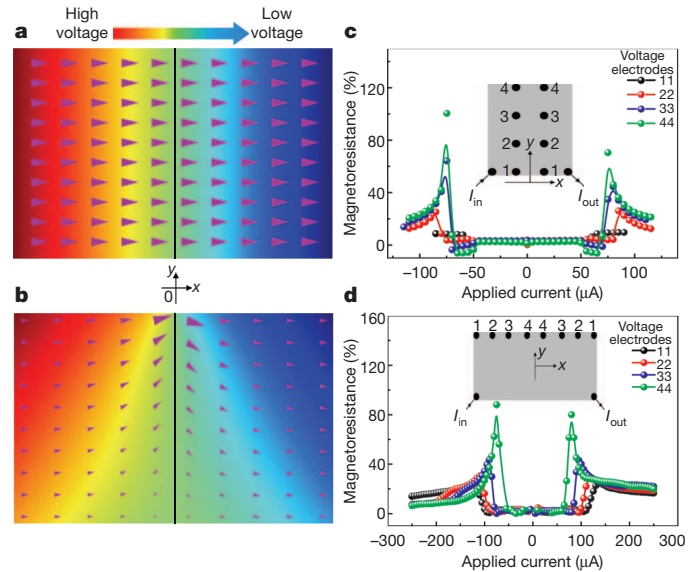


Figure 3 | Potential (colour scale) and current density (arrows) distributions and the effect of electrode geometry on magnetoresistance. The solid vertical lines in **a** and **b** show the p–n boundary. **a**, $B = 0$. The potential was symmetric about the x axis and current flowed along the x axis. **b**, $B = \mu^{-1}$. A positive Hall voltage appeared in the p region and a negative Hall voltage appeared in the n region. The boundary acted as a magnetic scattering resource. The current trajectory was distorted upward. **c**, As y_0 increased, the apparent maximum magnetoresistance of sample 18 increased. **d**, As x_0 decreased, the apparent maximum magnetoresistance of sample 26 increased. Insets in **c** and **d** show the placement of voltage contacts.

positive and negative Hall voltages built up in the minority and majority regions, respectively (Fig. 3b). In this situation, carriers were subjected to three fields: the electric field E_x along the x axis, the Lorentz force $q\mathbf{v}B$ and the Hall electric field E_y along the y axis where q is the elementary charge and \mathbf{v} is the drifting velocity of carriers. Far from the p–n boundary, $\mathbf{v}B = E_y$ ($|x| \gg 0$), carriers moved forward along the x axis. However, the magnitude of the Hall voltage near the boundary decreased to satisfy the voltage continuity condition. Thus when $E_y(|x| \approx 0) < E_y(|x| \gg 0) = \mathbf{v}B$, carriers were pushed by the net Lorentz force so that they no longer moved perpendicularly to the boundary. The boundary acted as a magnetic scattering resource and distorted the current trajectory. We ascribe the IMR we observed mainly to this p–n boundary. This mechanism for producing IMR is different from the mechanisms proposed in refs 8 and 21, which ascribe the magnetoresistances observed therein to shrinkage of the wavefunction of impurity states.

According to our finite-element-modelling simulation, current distortion occurred mainly within the region where $E_x^H \equiv dV_H/dx \geq E_x^e$, where V_H , E_x^H and E_x^e are Hall voltage, the Hall voltage gradient and the external electric field along the x axis, respectively. The degree of current distortion was estimated as $E_x^H/E_x^e \approx (\mu B)(y_0/x_0)$ where x_0 and y_0 are the x and y positions of the voltage electrodes (see Supplementary Information). Thus a larger apparent IMR could be acquired by placing voltage electrodes in suitable geometries. The apparent maximum IMR at 1.2 T of sample 18 rose from 10% to 80% with increasing y_0 (Fig. 3c), while that of sample 26 rose from 25% to 90% with decreasing x_0 (Fig. 3d), proving that our proposed physical scenario for IMR was reasonable.

On the basis of our finite-element-modelling prediction that the apparent IMR could be further enhanced by a large y_0/x_0 ratio, we designed an IMR device (sample 40) with a large W/L ratio of 50 (Fig. 4a). Like sample 20, sample 40 showed three kinds of field dependence of IMR (Fig. 4b). Besides a normal magnetoresistance at $I < 120 \mu A$, we also observed a crossover from the normal magnetoresistance to IMR at $120 \mu A \leq I < 250 \mu A$. With increasing magnetic

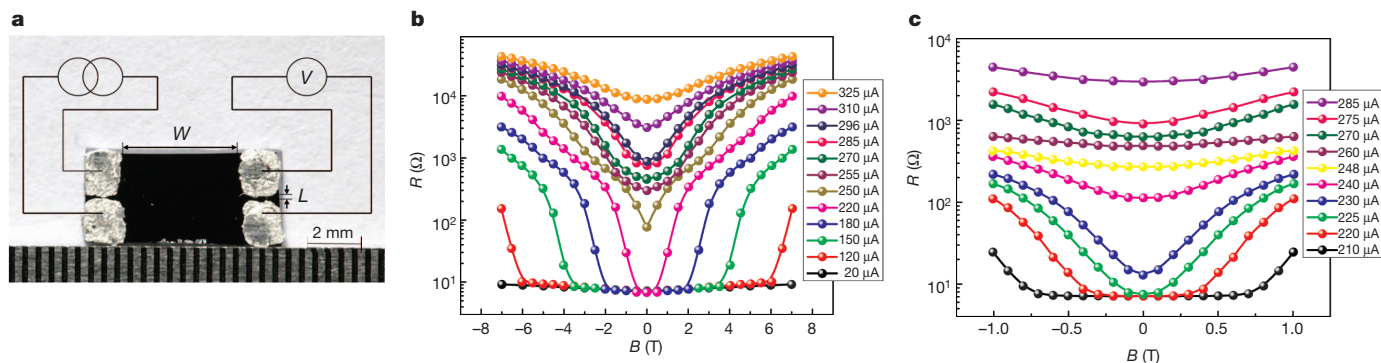


Figure 4 | Magnetoresistance of sample 40 at 300 K. **a**, The device geometry of sample 40. L is the spacing between two voltage electrodes and W is the spacing between current-voltage electrodes. The effect of magnetic field on magnetoresistance in sample 40 under different current conditions: **b**, 0–7 T; **c**, 0–1 T.

field, the resistance at 220 μA at first increased proportional to $(\mu_0 B)^2$ below a critical field $B_0 = 0.4$ T, above which the resistance leaped by three orders of magnitude. The IMR at 7 T and 220 μA reached $1.5 \times 10^5\%$, the largest magnetoresistance reported for silicon to the best of our knowledge. More interestingly, low-field sensitivity could be greatly enhanced by shifting B_0 from 6.0 T at 120 μA to about 0.1 T at 225 μA . The degree of current distortion was evaluated using the ratio E_x^H/E_x^e , which was also proportional to the geometry factor W/L . Therefore both the field sensitivity and magnitude of IMR could be further enhanced by a large W/L ratio. The resistance at 225 μA increased by 1,000% at 0.5 T, 100% at 0.2 T and 10% at 0.07 T (Fig. 4c). We tested the magnitude of this IMR effect and found it to be reproducible within approximately 10% given a fixed large W/L ratio (see Supplementary Information). By further increasing the W/L ratio, we achieved even larger giant and even more sensitive magnetoresistance. Above the transition region at $I \geq 250 \mu\text{A}$, the magnitude of the IMR at 7 T decreased gradually. At this region, the p - n boundary had been driven away from the voltage electrodes. However, like the IMR values reported in refs 1 and 7, our IMR could still be enhanced by varying the density of holes. No trace of saturation was observed up to 7 T.

Although IMR devices in non-magnetic materials^{1,3–8} exhibit giant magnitude at high magnetic field, few show low-magnetic-field performance as sensitive as that of their magnetic counterparts: giant-magnetoresistance²² or tunnel-magnetoresistance²³ devices. This hampers their application in magnetic readers or speed monitors. Our IMR device exhibits field sensitivity (evaluated as magnetoresistance divided by magnetic field) of 10% IMR at 0.07 T and 100% IMR at 0.2 T, which is nearly comparable with giant-magnetoresistance devices^{19,20}, suggesting an alternative device for the low-magnetic-field sensing industry. In addition, our device can operate at very high magnetic fields of several Tesla, at which conventional giant-magnetoresistance devices^{19,20,22} or organic magnetoresistance devices^{24,25} have already saturated and thus do not work. The versatility of our IMR device under both high and low magnetic fields suggest its application in wide-range sensing and could enable the integration of low- and high-magnetic-field sensing in a single device.

More importantly, our IMR device is based on semiconductor silicon, making it compatible with mature silicon technology. It could therefore easily be integrated into silicon chips, which is not feasible for its magnetic counterparts. Our IMR device can be modulated not only by electric field but also by magnetic field, so it should provide present electronics with a new mode of operation and more functionality based on the interplay between electronics and magnetic response.

METHODS SUMMARY

Two types of n -type silicon wafers were selected for fabricating our IMR devices. They were both (100) orientated and had the same thickness of 500 μm , but different resistivities of 30 Ωm and 10 Ωm . We used two main geometries: the first geometry is shown in Fig. 1a, where current is injected from the outer electrodes,

two electrodes along the x axis in the middle of the sample were used to measure voltage as well as resistance, and the other two electrodes along the y axis were used to measure the Hall voltage (van der Pauw method). The second geometry is shown in Fig. 4a, where four electrodes were placed on the corners of a rectangular sample (four-terminal method). The sample sizes were $W \times L$. Two other geometries with different electrode distributions (Fig. 3c and d) were also used to investigate the effect of geometry on IMR. An indium (>99.99%) electrode was used to make contacts with silicon. The electro- and magneto-transport properties were all measured using the four-terminal method. The magnetic field was perpendicular to the sample plane. To cancel out the influence of Hall voltage on resistance, the even part of the resistance, $[R(B) + R(-B)]/2$, was separated from the raw resistance data, and the odd part, $[R(B) - R(-B)]/2$, was eliminated (see the Supplementary Information). Except where otherwise noted, all the measurements were conducted at 300 K (see the Supplementary Information for the temperature dependence of IMR) and the resistivity of samples was 30 Ωm . Minority (hole) lifetimes were measured to be 100–200 μs for 30- Ωm samples. The dependence of IMR on resistivity is shown in the Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 April; accepted 18 July 2011.

- Delmo, M. P., Yamamoto, S., Kaisa, S., Ono, T. & Kobayashi, K. Large positive magnetoresistive effect in silicon induced by the space-charge effect. *Nature* **457**, 1112–1115 (2009).
- Solin, S. A., Thio, T., Hines, D. R. & Heremans, J. J. Enhanced room-temperature geometric magnetoresistance in inhomogeneous narrow-gap semiconductors. *Science* **289**, 1530–1532 (2000).
- Xu, R. *et al.* Large magnetoresistance in non-magnetic silver chalcogenides. *Nature* **390**, 57–60 (1997).
- Husmann, A. *et al.* Megagauss sensors. *Nature* **417**, 421–424 (2002).
- Hu, J. S. & Rosenbaum, T. F. Classical and quantum routes to linear magnetoresistance. *Nature Mater.* **7**, 697–700 (2008).
- Wu, L. H. *et al.* Room-temperature nonsaturating magnetoresistance of intrinsic bulk silicon in high pulsed magnetic fields. *Appl. Phys. Lett.* **98**, 112113 (2011).
- Delmo, M. P., Kaisa, S., Kobayashi, K. & Ono, T. Current-controlled magnetoresistance in silicon in non-Ohmic transport regimes. *Appl. Phys. Lett.* **95**, 132106 (2009).
- Schoonus, J. J. H. M., Haazen, P. P. J., Swagten, H. J. M. & Koopmans, B. Unraveling the mechanism of large room-temperature magnetoresistance in silicon. *J. Phys. D* **42**, 185011 (2009).
- Parish, M. M. & Littlewood, P. B. Non-saturating magnetoresistance in heavily disorder semiconductors. *Nature* **426**, 162–165 (2003).
- Hu, J. S., Parish, M. M. & Rosenbaum, T. F. Nonsaturating magnetoresistance of inhomogeneous conductors: comparison of experiment and simulation. *Phys. Rev. B* **75**, 214203 (2007).
- Parish, M. M. & Littlewood, P. B. Classical magnetotransport of inhomogeneous conductors. *Phys. Rev. B* **72**, 094417 (2005).
- Herring, C. Effects of random inhomogeneities on electrical and galvanomagnetic measurements. *J. Appl. Phys.* **31**, 1939–1953 (1960).
- Stroud, D. Generalized effective-medium approach to the conductivity of an inhomogeneous material. *Phys. Rev. B* **12**, 3368–3373 (1975).
- Stroud, D. & Pan, F. P. Effect of isolated inhomogeneities on the galvanomagnetic properties of solids. *Phys. Rev. B* **13**, 1434–1438 (1976).
- Stroud, D. & Pan, F. P. Magnetoresistance and Hall coefficient of inhomogeneous metals. *Phys. Rev. B* **20**, 455–465 (1979).
- Stroud, D. New exact results for the Hall coefficient and magnetoresistance of inhomogeneous two-dimensional metals. *Phys. Rev. B* **30**, 447–449 (1984).
- Guttal, V. & Stroud, D. Model for a macroscopically disordered conductor with an exactly linear high-field magnetoresistance. *Phys. Rev. B* **71**, 201304(R) (2005).

18. Magier, R. & Bergman, D. J. Strong-field magnetotransport of two-phase disordered media in two and three dimensions: Exact and approximate results. *Phys. Rev. B* **74**, 094423 (2006).
19. Duvail, J. L. *et al.* Giant magnetoresistance in hybrid nanostructures. *J. Magn. Magn. Mater.* **151**, 324–332 (1995).
20. Parkin, S. S. P. Giant magnetoresistance in magnetic nanostructures. *Annu. Rev. Mater. Sci.* **25**, 357–388 (1995).
21. Schoonus, J. J. H. M., Bloom, F. L., Wagemans, W., Swagten, H. J. M. & Koopmans, B. Extremely large magnetoresistance in boron-doped silicon. *Phys. Rev. Lett.* **100**, 127202 (2008).
22. Baibich, M. N. *et al.* Giant magnetoresistance in Fe(001)/Cr(001) superlattices. *Phys. Rev. Lett.* **61**, 2472–2475 (1988).
23. Parkin, S. S. P. *et al.* Giant tunneling magnetoresistance at room temperature with MgO (100) tunnel barrier. *Nature Mater.* **3**, 862–867 (2004).
24. Hu, B. & Wu, Y. Tuning magnetoresistance between positive and negative in organic semiconductors. *Nature Mater.* **6**, 985–991 (2007).
25. Bloom, F. L., Wagemans, W., Kemerink, M. & Koopmans, B. Separating positive and negative magnetoresistance in organic semiconductor devices. *Phys. Rev. Lett.* **99**, 257201 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is supported by the Ministry of Science and Technology of China (grant 2009CB929202) and the National Science Foundation of China (grants 11074141 and U0734001). We used the resources of the Beijing National Center for Electron Microscopy.

Author Contributions C.H.W. designed research blueprints, did the main experimental measurements and the finite element modelling and wrote the manuscript. X.Z.Z. contributed to project design, experimental and theoretical analysis, manuscript writing and whole project supervision. X.L.G. and J.M.W. contributed to the measurements. X.Y.T. contributed to the modelling work. All authors contributed to result analysis and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to X.Z.Z. (xzzhang@tsinghua.edu.cn).

METHODS

Silicon types. Besides the two silicon wafers with resistivities of $30\ \Omega\text{ m}$ and $10\ \Omega\text{ m}$, we also used two other kinds of silicon (from Tianjin Far Hongda Electronics) with the same orientation and the same thickness but different resistivities of $4.0\ \Omega\text{ m}$ and $1.0\ \Omega\text{ m}$ to investigate the dependence of magnetoresistance on resistivity (the results are in the Supplementary Information). The highest-resistivity ($30\ \Omega\text{ m}$) silicon was fabricated using floating zone techniques. The hole lifetime of the $30\text{-}\Omega\text{ m}$ samples was $100\text{--}200\ \mu\text{s}$. The dependence of IMR on resistivity is given in the Supplementary Information.

Electrode properties. Indium electrodes $0.5\text{--}2\text{ mm}$ in size were cold-pressed onto the surfaces of the samples. The $\text{In}/\text{SiO}_2/\text{Si}$ electrode was not a perfect ohmic contact; instead, it had a small Schottky barrier of about 0.2 eV , which reduced the magnetoresistance of our device at low temperature. However, the intrinsic SiO_2 layer (1.6 nm thick) dominated the resistance of the electrodes at room temperature, which helped to inject holes from indium to silicon (see Supplementary Information).

Measurement setup. We used a Keithley 2400 sourcemeter. All the measurements were conducted using the Magnetic Property Measurement System (MPMS XL-7,

Quantum Design). Two typical geometries are shown in Fig. 2a and Fig. 4a. Hall coefficient and transport properties were measured with the Van der Pauw method and the four-terminal method, respectively. The magnetic field was kept perpendicular to the sample plane. The temperature dependence of IMR was measured as shown in the Supplementary Information. To cancel out the influence of Hall voltage on resistance, the even part of the resistance, $[R(B) + R(-B)]/2$, was separated from the raw resistance data, and the odd part, $[R(B) - R(-B)]/2$, was eliminated (see the Supplementary Information).

Finite element modelling. We used the method proposed in refs 9–11 to model the potential and current density distributions in the samples. According to this method, we separated the samples into pieces. Each piece had its own carrier type, density and mobility. Here we also separated a rectangular sample into holes and electron regions with a p–n boundary between them (Fig. 3). For simplicity, the carrier density and the absolute value of the mobility were set to be the same for both regions, $n = p$ and $|\mu_h| = |\mu_e| = |\mu|$. The conductivity tensor σ of both regions was defined as $\sigma_{xx} = \sigma_{yy} = \sigma_0/[1 + (\mu B)^2]$, $\sigma_{xy} = -\sigma_{yx} = -\mu B\sigma_0/[1 + (\mu B)^2]$. At $B = 0$ the conductivity $\sigma_0 = ne\mu$. Current I was injected and collected at the left and right boundaries of the sample, respectively; other boundaries were set to be insulating.

Overcoming lability of extremely long alkane carbon–carbon bonds through dispersion forces

Peter R. Schreiner¹, Lesya V. Chernish², Pavel A. Gunchenko², Evgeniya Yu. Tikhonchuk², Heike Hausmann¹, Michael Serafin³, Sabine Schlecht³, Jeremy E. P. Dahl⁴, Robert M. K. Carlson⁴ & Andrey A. Fokin^{1,2}

Steric effects in chemistry are a consequence of the space required to accommodate the atoms and groups within a molecule, and are often thought to be dominated by repulsive forces arising from overlapping electron densities (Pauli repulsion). An appreciation of attractive interactions such as van der Waals forces (which include London dispersion forces) is necessary to understand chemical bonding and reactivity fully. This is evident from, for example, the strongly debated origin of the higher stability of branched alkanes relative to linear alkanes^{1,2} and the possibility of constructing hydrocarbons with extraordinarily long C–C single bonds through steric crowding³. Although empirical bond distance/bond strength relationships have been established for C–C bonds⁴ (longer C–C bonds have smaller bond dissociation energies), these have no present theoretical basis⁵. Nevertheless, these empirical considerations are fundamental to structural and energetic evaluations in chemistry^{6,7}, as summarized by Pauling⁸ as early as 1960 and confirmed more recently⁴. Here we report the preparation of hydrocarbons with extremely long C–C bonds (up to 1.704 Å), the longest such bonds observed so far in alkanes. The prepared compounds are unexpectedly stable—noticeable decomposition occurs only above 200 °C. We prepared the alkanes by coupling nanometre-sized, diamond-like, highly rigid structures known as diamondoids⁹. The extraordinary stability of the coupling products is due to overall attractive dispersion interactions between the intramolecular H···H contact surfaces, as is evident from density functional theory computations with¹⁰ and without inclusion of dispersion corrections.

“Matter will always display attraction” was J. D. van der Waals’ favourite maxim¹¹, but this precept seems to have been partly forgotten. Literally stretching the limits of chemical bonding improves our understanding of the nature of stereoelectronic effects and the relative weights of covalent contributions relative to noncovalent contributions. General consensus exists regarding correlations between C–C bond lengths and their bond dissociation energies (BDEs) for a broad range of strained and unstrained compounds: shorter bonds are considered stronger, and vice versa. However, there are many exceptions to this relationship for bonds between elements other than carbon, emphasizing that there is no generalizable physical basis for this assumption⁵. Although practically all of these exceptions rely on the incorporation of highly electronegative atoms, we show here that alkanes with the longest C–C single bonds ever observed can still be quite stable. Such compounds can be realized by shifting the energy balance in favour of attractive dispersion interactions that outweigh to a large degree the repulsive dispersion contributions leading to C–C bond elongation. Our findings have consequences for understanding rotational barriers and thermodynamic preferences of branched alkanes over linear alkanes^{1,2,12}, and for the design of structures using attractive dispersion interactions. The examination of model systems to probe, rigorously understand and eventually control such ‘hydrophobic interactions’ (a term used in the life sciences) is key to advancing many aspects of molecular recognition.

The general recipe for elongating chemical bonds involves steric crowding¹³. This approach works well for, for example, structures 1–4 (Fig. 1), which have remarkably long C–C bonds, of up to 1.72 Å (C–C bond lengths of up to 1.78 Å have been reported for silicon-containing structures¹⁴), but reaches its limit of applicability with the highly crowded ‘classic riddle’¹⁵ hexaphenyl ethane (5, R = H), which has not yet been realized because its BDE apparently is too small (computed to be 17 kcal mol^{−1}; ref. 16) to allow its isolation. A bond length of 1.67 Å was determined experimentally¹⁷ for persistent—yet sterically much more crowded—hexakis(3,5-di-*t*-butylphenyl)ethane (5, R = *t*-butyl (*t*-Bu)). Such low BDEs also result from benzylic resonance stabilization of the product triphenylmethyl radicals, whose formation is suppressed by holding the fragments in place through molecular bridges in the related compounds 3 and 4¹⁸.

The most sterically crowded alkanes prepared so far are 1¹⁹ and 2²⁰, with C–C bonds of up to 1.65 Å. They are considered thermally labile, as expressed in the half-life of 1 h for 2 at 167 °C. Even more sterically crowded alkanes were deemed impossible because the BDEs for C–C bonds longer than 1.65 Å were empirically estimated to be around only 41 kcal mol^{−1} (refs 3, 4).

Our strategy to overcome the overall bond weakening through repulsive interactions is to design structures that additionally feature attractive dispersion interactions, by controlling the number and lengths of hydrogen–hydrogen van der Waals contacts surrounding each C–C bond under consideration. This idea is inconsistent with the general assumption that strained alkane C–C bonds are dominated by repulsive interactions²¹, and we will use quantum chemical computations to show that the balance between repulsive and attractive van der

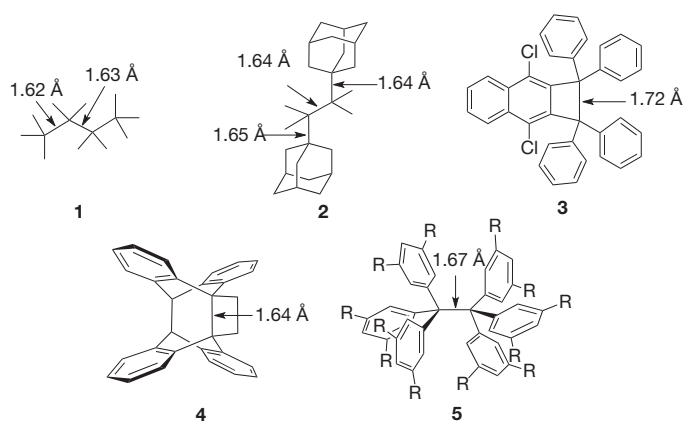


Figure 1 | Hydrocarbons with exceptionally long C–C bonds. Structures 1–4 have been reported experimentally; structure 5 (R = H) has not been observed but 5 (R = *t*-Bu) is experimentally known (bond length given). All of these structures are thermally labile and have half-lives of only a few hours upon moderate heating.

¹Institut für Organische Chemie der Justus-Liebig-Universität, Heinrich-Buff-Ring 58, D-35392 Giessen, Germany. ²Department of Organic Chemistry, Kiev Polytechnic Institute, 37 Pobeda Avenue, Kiev 03056, Ukraine. ³Institut für Anorganische Chemie der Justus-Liebig-Universität, Heinrich-Buff-Ring 58, D-35392 Giessen, Germany. ⁴Stanford University, Stanford Institute for Materials & Energy Science, 476 Lomita Mall, Stanford, California 94305, USA.

Waals interactions can be shifted significantly. We used this strategy to prepare alkanes with unprecedented thermal stability and C–C bond lengths. A second aspect of our strategy is that the radicals formed from C–C bond dissociation should be structurally very similar to the hydrocarbon moiety in the undissociated starting material, to avoid stabilization through interactions only present in the relaxed alkyl radical structures²².

These considerations led us to the design and preparation of coupled diamondoid molecules because these have large, hydrogen-terminated contact areas (Fig. 2) and their corresponding tertiary radicals are structurally very similar to their hydrocarbon precursors, such that radical stabilization through geometrical relaxation is minimized. Diamondoids are nanometre-sized (0.4–1.2 nm), hydrogen-terminated, diamond-like alkanes that are available through synthesis (only for the smallest members of this family) or through isolation from petroleum⁹. These true nanodiamonds, of which adamantane is the smallest, consist of a series formed by adding adamantane subunits to a tetrahedral C₁₀H₁₆ core; the naming of the simplest nanodiamonds follows from the number of adamantane moieties⁹ (diamantane, triamantane and so forth; Fig. 2).

The tertiary diamondoid bromides of hydrocarbons 6–8 (Fig. 2) readily undergo Wurtz coupling at 145 °C in xylene to give the heterodimers 7•7 (65%), 6•8 (25%) and 7•8 (21%) (where the point denotes the C–C bond), which were chemically fully characterized. All diamondoid adducts crystallize well and were subjected to X-ray analysis (Fig. 2), revealing extraordinarily long central C–C bonds (1.647–1.704 Å).

Notably, all three compounds have high melting points, and we assessed their thermal stability using differential scanning calorimetry (DSC) and thermogravimetric analyses (TGA). These analyses reveal that 7•7 is stable up to at least 300 °C and melts at about 360 °C. Similarly, 6•8 (melting point, 310 °C) slowly decomposes above 300 °C, whereas 7•8 begins decomposition at only 220 °C. The monitoring of the TGA experiments with a mass detector revealed that the volatile decomposition products are the parent hydrocarbons 6 and 7. Such high melting points are typical for stable cage hydrocarbons such as the diamondoids (melting points: 6, 270 °C; 7, 244 °C; 8, 225 °C), by marked contrast with the low melting points of conformationally flexible alkanes²³. To quantify further the great stability of 7•8, we determined the hydrogen transfer reaction energies in the presence of 9,10-dihydroanthracene²⁰ as a hydrogen donor in pressure-tight steel containers (for details, see Supplementary Figs 8 and 9 and Supplementary Table 1), finding a reaction enthalpy of –24.6 kcal mol^{–1} in the range of 190–210 °C. As the temperature for this hydrogen transfer reaction and the onset of decomposition of 7•8 are close, the activation free energy associated with this reaction ($\Delta G_{573}^\ddagger = 37.5 \pm 7.9$ kcal mol^{–1}) must be attributed to the interfragment

C–C bond breaking in 7•8. This is in line with the expectation that 7•8 is more stable than 2 ($\Delta G_{573}^\ddagger = 30.5$ kcal mol^{–1}), despite its considerably longer central C–C bond length.

The gas-phase stability of the heterodimers was also assessed by gas chromatography mass spectrometry measurements. For instance, 7•7 can readily be identified by its molecular ion mass peak, even at an inlet temperature of 280 °C and a retention time of 120 min.

From known C–C bond distance/bond energy correlations⁴, these heterodimers are all expected to be thermally unstable. Hence, the lability of their central C–C bonds must be energetically overcompensated for by favourable bonding interactions. We therefore determined which interactions are responsible for the stabilities of 7•7, 6•8 and 7•8. The X-ray crystal structure analyses reveal that the lengths of the H···H contacts between the two hydrocarbon moieties are in the range of 1.9–2.6 Å, with the majority being around 2.2–2.3 Å. This corresponds well to the optimal H···H distances found for molecular crystals of many organic structures (2.2–2.4 Å; ref. 24); the H···H contacts in the adamantane X-ray crystal structure are 2.37–2.46 Å in length²⁵.

We also performed a computational analysis of 7•7, in the same symmetry (C₂) as found in the X-ray structure, and of 5 using various density functional theory (DFT) approaches. As the standard implementation of DFT does not explicitly include dispersion interactions (for example in the popular B3LYP functional combination), this allows an analysis of the results by comparison with dispersion-corrected (DFT-D) levels of theory¹⁰ (B3LYP-D). These results are compared with modern functionals that have been extensively reparameterized (for example M06; Table 1). To validate our computational approach, we computed the reaction enthalpy of the hydrogen transfer from 9,10-dihydroanthracene and found that our reference computations at B3LYP-D/6-31G(d,p) give –26.9 kcal mol^{–1} (at 200 °C), in excellent agreement with experiment (–24.6 kcal mol^{–1}), only on inclusion of dispersion interactions. Similar results were obtained with a modern functional (M06-2X; ref. 26) and another (B97D) that more properly account (to different degrees) for dispersion interactions (Supplementary Table 1). Neglect of dispersion, as in uncorrected B3LYP/6-31G(d,p), gives an error of nearly 30 kcal mol^{–1}.

The B3LYP-D and M06-2X approaches reproduce the central C–C bond distances quite well (an exact match cannot be expected because computed values inevitably differ slightly from the experimental data owing to approximations in the DFT formulations and the differences arising from gas-phase versus condensed-phase structures), lending credibility to the computations. For 7•7, the inclusion of dispersion corrections increases the BDE significantly. A BDE of 71 kcal mol^{–1} for 7•7, which is nearly 30 kcal mol^{–1} above the expected empirical value⁴, is in agreement with the experimentally found high stability.

For 5 (R = H), inclusion of dispersion corrections reduces the BDE and increases the central C–C bond distance, whereas it significantly decreases that in 7•7. This indicates the overall dominance of repulsive interactions between the neighbouring phenyl rings in 5 (R = H). Although the phenyl rings in 5 (R = H) have a favourable, distorted T-shape benzene dimer²⁷ orientation relative to each other, their attractive dispersion interactions are insufficient to allow the preparation and isolation of 5 (R = H) at ambient temperatures. Remarkably, the addition of all-*meta* *t*-Bu groups to give 5 (R = *t*-Bu) increases the BDE (relative to 5 (R = H)) and decreases the central C–C bond distance. Indeed, 5 (R = *t*-Bu) has been fully characterized by crystal structure analysis¹⁷. By contrast with 5 (R = H), the inclusion of dispersion corrections decreases the central C–C bond distance for 5 (R = *t*-Bu), as it does for 7•7: This must be a consequence of attractive dispersion interactions resulting from addition of the *t*-Bu groups. Again, the H···H contact distances of the *t*-Bu groups in 5 (R = *t*-Bu) are around 2.1–2.5 Å, which is comparable to our heterodimer structures, demonstrating the similarities in the sources of their stabilization.

The notion of attractive rather than repulsive H···H contacts touches on many aspects of chemistry, biology and the materials sciences. For

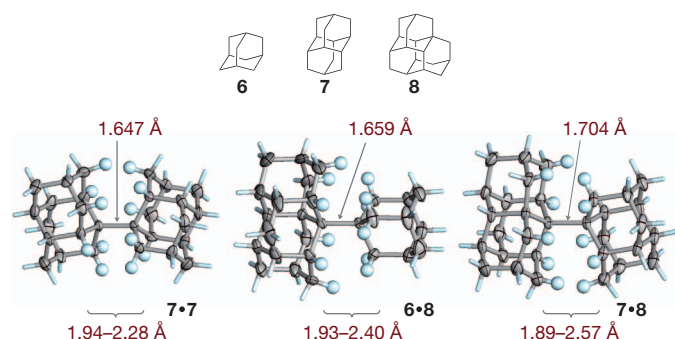
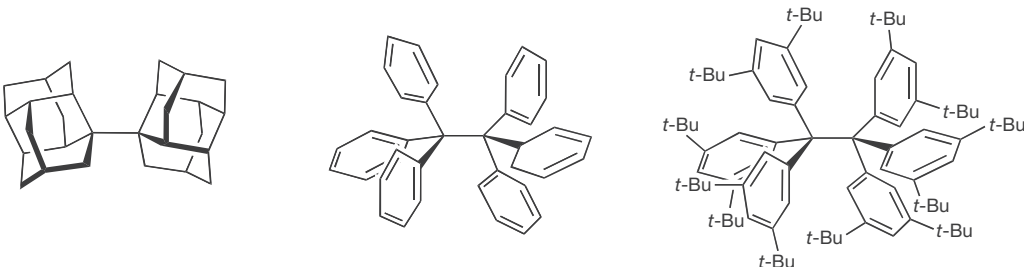


Figure 2 | Diamondoids and X-ray crystal structures of their coupling products with very long central C–C bonds. Adamantane (6), diamantane (7) and triamantane (8) as nanodiamond building blocks for their coupling products diamantane–diamantane (7•7), adamantane–triamantane (6•8) and diamantane–triamantane (7•8). The hydrogen-terminated surfaces are shown with the hydrogen atoms in light blue; the curly brackets indicate the distance ranges for the H···H contacts around the central C–C bonds.

Table 1 | The BDEs and C–C bond lengths of **7·7 and **5** computed at various levels of DFT**


Method/quantity	BDE (kcal mol ⁻¹)	C–C (Å)	BDE (kcal mol ⁻¹)	C–C (Å)	BDE (kcal mol ⁻¹)	C–C (Å)
B3LYP/6-31G(d,p)	43.9	1.674	–20.9	1.730	–26.1	1.709
B3LYP-D/6-31G(d,p)	70.7	1.653	10.3	1.735	44.5	1.674
B97D/6-31G(d,p)	64.5	1.668	6.5	1.791	38.8	1.698
M06-2X/6-31G(d,p)	65.8	1.648	12.3	1.702	33.0	1.669
Experiment	—	1.647	—	—	—	1.670(3)

The experimental bond distance for **5** (R = *t*-Bu) is from ref. 17. Structures are drawn for best possible visibility. 'D' denotes a correction for dispersion energies; the M06-2X functional has been extensively reparameterized to include some amount of dispersion. The parenthetical uncertainty in the **5** (R = *t*-Bu) C–C bond length is the standard deviation in X-ray standard notation.

instance, protobranching, defined as net attractive 1,3-alkyl-alkyl stabilizing interactions, has been suggested (and criticized¹²) to be responsible for the higher thermodynamic stability of branched alkanes over unbranched alkanes². It is likely that these overall stabilizing interactions receive large contributions from favourable H···H contacts. Another example is the 'corset effect', whereby apparent steric crowding around labile molecular moieties stabilizes the overall structure kinetically; a prime example is the preparation and isolation of tetra-*t*-butyltetrahedrane²⁸ (Supplementary Table 2; the less crowded parent hydrocarbon is yet unknown). This stabilization can alternatively be interpreted as arising from favourable van der Waals contacts of the *t*-Bu groups; this suggestion is supported by the value of -3.1 kcal mol⁻¹ computed for the isodesmic equation 2 di-*t*-butyltetrahedrane → tetra-*t*-butyltetrahedrane + tetrahedrane at B3LYP-D/6-31G(d,p). Along these lines, it is notable that many recently discovered carbene-stabilized complexes also involve bulky pendant alkyl groups²⁹ that may contribute to the overall stabilization of these otherwise labile systems³⁰.

METHODS SUMMARY

The diamondoid heterodimers were prepared by refluxing the respective bromodiamondoid precursors in a small volume of dry *m*-xylene under argon atmosphere in the presence of sodium metal. After work-up and compound separation on silica gel, the final products were crystallized from *n*-hexane. Experimental details and compound characterizations are described in detail in Methods and Supplementary Information. DSC measurements were performed in platinum–corundum double-layer crucibles under argon. TGA analyses (coupled with a mass spectrometer) were conducted in corundum crucibles under argon atmosphere. All thermal analyses were temperature-calibrated. The X-ray crystallographic data for **7·7**, **6·8** and **7·8** were collected at 193 K using molybdenum K α radiation and a graphite monochromator. The structures were solved by direct methods and refined by using full-matrix least-squares analyses; all non-hydrogen atoms were treated anisotropically. Importantly, all hydrogen atoms could be found in the difference Fourier syntheses and were refined isotropically.

The electronic structure computations were carried out with the Gaussian03 and Gaussian09 program suites. All structures were fully optimized and characterized as minima (by computing analytical second derivatives) of their respective potential energy hypersurfaces at the levels of DFT given in the text. All optimized geometries (*x*–*y*–*z* coordinates) and absolute electronic energies are given in Supplementary Table 3.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 January; accepted 15 July 2011.

- Grimme, S. Seemingly simple stereoelectronic effects in alkane isomers and the implications for Kohn–Sham density functional theory. *Angew. Chem. Int. Ed.* **45**, 4460–4464 (2006).
- Wodrich, M. D. *et al.* The concept of protobranching and its many paradigm shifting implications for energy evaluations. *Chem. Eur. J.* **13**, 7731–7744 (2007).

- de Silva, K. M. N. & Goodman, J. M. What is the smallest saturated acyclic alkane that cannot be made? *J. Chem. Inf. Model.* **45**, 81–87 (2005).
- Zavitsas, A. A. The relation between bond lengths and dissociation energies of carbon-carbon bonds. *J. Phys. Chem. A* **107**, 897–898 (2003).
- Kaupp, M., Metz, B. & Stoll, H. Breakdown of bond length-bond strength correlation: a case study. *Angew. Chem. Int. Ed.* **39**, 4607–4609 (2000).
- Gordy, W. A relation between bond force constants, bond orders, bond lengths, and the electronegativities of the bonded atoms. *J. Chem. Phys.* **14**, 305–320 (1946).
- Huggins, M. L. Atomic radii. IV. Dependence of interatomic distance on bond energy. *J. Am. Chem. Soc.* **75**, 4126–4133 (1953).
- Pauling, L. *The Nature of the Chemical Bond* (Cornell Univ. Press, 1960).
- Schwertfeger, H., Fokin, A. A. & Schreiner, P. R. Diamonds are a chemist's best friend: diamondoid chemistry beyond adamantane. *Angew. Chem. Int. Ed.* **47**, 1022–1036 (2008).
- Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **25**, 1463–1473 (2004).
- Tang, K.-T. & Toennies, J. P. Johannes Diderik van der Waals: a pioneer in the molecular sciences and Nobel prize winner in 1910. *Angew. Chem. Int. Ed.* **49**, 9574–9579 (2010).
- Gronert, S. The folly of protobranching: turning repulsive interactions into attractive ones and rewriting the strain/stabilization energies of organic chemistry. *Chem. Eur. J.* **15**, 5372–5382 (2009).
- Suzuki, T., Takeda, T., Kawai, H. & Fujiwara, K. Ultralong C–C bonds in hexaphenylethane derivatives. *Pure Appl. Chem.* **80**, 547–553 (2008).
- Fritz, G., Wartanessian, S., Matern, E., Höhle, W. & von Schnering, H. G. Bildung siliciumorganischer Verbindungen. 85. Bildung, Reaktionen und Struktur des 1,1,3,3-tetramethyl-2,4-bis(trimethylsilyl)-1,3-disilabicyclo[1.1.0]butans. *Z. Allg. Anorg. Chem.* **475**, 87–108 (1981).
- McBride, J. M. The hexaphenylethane riddle. *Tetrahedron* **30**, 2009–2022 (1974).
- Vreven, T. & Morokuma, K. Prediction of the dissociation energy of hexaphenylethane using the ONIOM(MO: MO: MO) method. *J. Phys. Chem. A* **106**, 6167–6170 (2002).
- Kahr, B., van Engen, D. & Mislow, K. Length of the ethane bond in hexaphenylethane and its derivatives. *J. Am. Chem. Soc.* **108**, 8305–8307 (1986).
- Kammermeier, S., Jones, P. G. & Herges, R. [2+2] cycloaddition products of tetrahydrodianthracene: experimental and theoretical proof of extraordinary long C–C single bonds. *Angew. Chem. Int. Ed. Engl.* **36**, 1757–1760 (1997).
- Winiker, R., Beckhaus, H. D. & Rüchardt, C. Thermische Stabilität, Spannungsenthalpie und Struktur symmetrisch hexaalkylierter Ethane. *Chem. Ber.* **113**, 3456–3476 (1980).
- Flamm-ter Meer, M. A., Beckhaus, H. D., Peters, K., von Schnering, H. G. & Rüchardt, C. Thermolabile hydrocarbons, XXVII. 2,3-di-1-adamantyl-2,3-dimethylbutane; long bonds and low thermal stability. *Chem. Ber.* **118**, 4665–4673 (1985).
- Rüchardt, C. & Beckhaus, H. D. Towards an understanding of the carbon-carbon bond. *Angew. Chem. Int. Ed. Engl.* **19**, 429–440 (1980).
- Rüchardt, C. & Beckhaus, H.-D. Consequences of strain for the structure of aliphatic molecules. *Angew. Chem. Int. Ed. Engl.* **24**, 529–538 (1985).
- Boese, R., Weiss, H.-C. & Bläser, D. The melting point alternation in the short-chain *n*-alkanes: single-crystal X-ray analyses of propane at 30 K and of *n*-butane to *n*-nonane at 90 K. *Angew. Chem. Int. Ed.* **38**, 988–992 (1999).
- Grimme, S. *et al.* When do interacting atoms form a chemical bond? Spectroscopic measurements and theoretical analyses of dideuteriophenanthrene. *Angew. Chem. Int. Ed.* **48**, 2592–2595 (2009).
- Donohue, J. & Goodman, S. H. The crystal structure of adamantane: an example of a false minimum in least squares. *Acta Crystallogr.* **22**, 352–354 (1967).
- Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120**, 215–241 (2008).

27. Takatani, T. & Sherrill, C. D. Performance of spin-component-scaled Moller-Plesset theory (SCS-MP2) for potential energy curves of noncovalent interactions. *Phys. Chem. Chem. Phys.* **9**, 6106–6114 (2007).
28. Maier, G., Pfromm, S., Schäfer, U. & Matusch, R. Tetra-*tert*-butyltetrahedrane. *Angew. Chem. Int. Edn Engl.* **17**, 520–521 (1978).
29. Wang, Y. Z. & Robinson, G. H. Unique homonuclear multiple bonding in main group compounds. *Chem. Commun. (Camb.)* 5201–5213 (2009).
30. Dyker, C. A. & Bertrand, G. Soluble allotropes of main-group elements. *Science* **321**, 1050–1051 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Grimme for providing an implementation of his dispersion correction technique and for discussions. We are grateful for support from the Deutsche Forschungsgemeinschaft and the National Science Foundation of the USA, and in part from the Department of Energy, Office of Basic Energy Sciences, Division of Materials Science and Engineering, under contract DE-AC02-76SF00515;

the Ministry of Science and Education of Ukraine; and the Ukrainian State Basic Research Fund.

Author Contributions P.R.S. and A.A.F. formulated the initial working hypothesis and provided, analysed and interpreted all experimental data. L.V.C., P.A.G. and E.Yu.T. carried out the coupling experiments. H.H. recorded and analysed all NMR data. M.S. solved all X-ray structures. S.S. provided and interpreted the DSC and TGA analyses. J.E.P.D. and R.M.K.C. isolated and purified the diamondoids. The manuscript was written by P.R.S. and A.A.F.

Author Information X-ray crystal structures have been deposited in the Cambridge Crystallographic Database under the deposition numbers CCDC 805315 (**7•7**), CCDC 806293 (**6•8**) and CCDC 806294 (**7•8**). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.R.S. (prs@org.chemie.uni-giessen.de) or A.A.F. (aaf@xtf.ntu-kpi.kiev.ua).

METHODS

Diamondoid coupling procedure. For the Wurtz coupling, 1 mmol of the chosen bromodiamondoid precursors was dissolved in a small volume of dry *m*-xylene and refluxed (140–150 °C in the oil bath) in a two-neck flask fitted with an argon inlet and an anchor stirrer with an air-cooled condenser under a slow stream of argon. Small pieces of sodium (0.3 g, or 13 mmol, in total) were added to the stirred reaction mixture over 1.5 h. After adding all of the sodium, the mixture was refluxed for a total of 4 h and cooled to 50 °C; then the excess of sodium was quenched with methanol. After cooling to room temperature (23 ± 2 °C), the reaction mixture was filtered and washed with water, evaporated and separated on silica gel (*n*-hexane); the final products were crystallized from *n*-hexane. All compounds were characterized by nuclear magnetic resonance spectroscopy, high-resolution mass spectrometry, elemental analysis and X-ray crystal structure determination.

Computations. All geometries were fully optimized at the stated level of theory, described using the standard abbreviations: B3, Becke's three-parameter exchange functional³¹; LYP, Lee–Yang–Parr correlation functional³²; D, dispersion correction according to refs 33, 34; M06-2X, Truhlar's high-nonlocality functional with double the amount of nonlocal exchange³⁵; B97, Becke's 1997 exchange-correlation functional³⁵. We used a standard 6-31G³⁶ basis set with polarization functions on carbon (d) and hydrogen (p). All structures were characterized as minima of their respective potential energy hypersurfaces by confirming that all computed harmonic vibrational frequencies are real. These frequencies were also used to derive zero-point vibrational energy corrections to the relative energies. We used the Gaussian03³⁷ (version D.02 for adding Grimme's dispersion correction to B3LYP) and the Gaussian09³⁸ programs (version B.01) for all computations.

Thermogravimetric and differential scanning calorimetric analyses. DSC measurements were performed in a Netzsch Pegasus 404 C calorimeter in platinum–corundum double-layer crucibles under an argon flow of 50 ml min^{−1} at a heating rate of 10 K min^{−1}. TGA analyses were conducted in a Netzsch Luxx STA 409 PC apparatus coupled to an Aëlos QMS 403 C mass spectrometer. The samples were

heated in corundum crucibles in an argon atmosphere at a heating rate of 10 K min^{−1}. Both instruments were temperature-calibrated in the range from room temperature to 1,100 °C with standard element samples of indium, tin, bismuth, aluminium, silver and gold. The results for the compounds under consideration here are graphically summarized in Supplementary Figs 1–7. The reasons for the differences between the behaviour seen in Supplementary Fig. 2 and that seen in Supplementary Figs 1 and 3 can be found in the construction geometry of the apparatus used in our work. The TGA and mass spectrometry units are connected by a heated transfer line that is held at a temperature of 250 °C and is about 1 m in length. The volatile adamantane released from **6•8** can easily pass through this transfer line and is therefore detected as it forms. The much less volatile diamantane produced from **7•7** and **7•8** has to pass through the transfer line in stepwise sublimation processes and only reaches the mass spectrometer when the temperature at the TGA unit has reached a value much higher than 250 °C. The maximum concentration of diamantane is found at 550 °C in both Supplementary Fig. 1 and Supplementary Fig. 3. This is consistent with the formation of diamantane in both cases.

31. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
32. Lee, C. T., Yang, W. T. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B* **37**, 785–789 (1988).
33. Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27**, 1787–1799 (2006).
34. Schwabe, T. & Grimme, S. Towards chemical accuracy for the thermodynamics of large molecules: new hybrid density functionals including non-local correlation effects. *Phys. Chem. Chem. Phys.* **8**, 4398–4401 (2006).
35. Becke, A. D. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.* **107**, 8554–8560 (1997).
36. Binkley, J. S., Pople, J. A. & Hehre, W. J. Split valence basis sets. *J. Am. Chem. Soc.* **102**, 939–947 (1980).
37. Frisch, M. J., et al. *Gaussian 03 v.D.02* (Gaussian, Inc., 2003).
38. Frisch, M. J., et al. *Gaussian09 v.B.01* (Gaussian, Inc., 2009).

Linking mantle plumes, large igneous provinces and environmental catastrophes

Stephan V. Sobolev^{1,2*}, Alexander V. Sobolev^{3,4,5*}, Dmitry V. Kuzmin^{4,6}, Nadezhda A. Krivolutsкая⁵, Alexey G. Petrunin^{1,2}, Nicholas T. Arndt³, Viktor A. Radko⁷ & Yuri R. Vasiliev⁶

Large igneous provinces (LIPs) are known for their rapid production of enormous volumes of magma (up to several million cubic kilometres in less than a million years)¹, for marked thinning of the lithosphere^{2,3}, often ending with a continental break-up, and for their links to global environmental catastrophes^{4,5}. Despite the importance of LIPs, controversy surrounds even the basic idea that they form through melting in the heads of thermal mantle plumes^{2,3,6–10}. The Permo-Triassic Siberian Traps¹¹—the type example and the largest continental LIP^{1,12}—is located on thick cratonic lithosphere^{1,12} and was synchronous with the largest known mass-extinction event¹. However, there is no evidence of pre-magmatic uplift or of a large lithospheric stretching⁷, as predicted above a plume head^{2,6,9}. Moreover, estimates of magmatic CO₂ degassing from the Siberian Traps are considered insufficient to trigger climatic crises^{13–15}, leading to the hypothesis that the release of thermogenic gases from the sediment pile caused the mass extinction^{15,16}. Here we present petrological evidence for a large amount (15 wt%) of dense recycled oceanic crust in the head of the plume and develop a thermomechanical model that predicts no pre-magmatic uplift and requires no lithospheric extension. The model implies extensive plume melting and heterogeneous erosion of the thick cratonic lithosphere over the course of a few hundred thousand years. The model suggests that massive degassing of CO₂ and HCl, mostly from the recycled crust in the plume head, could alone trigger a mass extinction and predicts it happening before the main volcanic phase, in agreement with stratigraphic and geochronological data for the Siberian Traps and other LIPs⁵.

Petrological studies of Siberian Traps and associated alkaline rocks reveal high temperatures (1,600–1,650 °C)^{14,17} in their mantle sources. Olivine compositions in samples from lower units of the Norilsk lava section provide evidence that the mantle source of the Siberian Traps was unusually rich in ancient recycled oceanic crust¹⁴, in agreement with earlier predictions¹⁰. For the main volcanic phase, however, such data were unavailable. Here we report 2,500 new olivine analyses and host-rock compositions for 45 basalts covering the main stages of tholeiitic magmatism in three key localities: the Norilsk area, the Putorana plateau and the Maymecha–Kotuy province (Fig. 1a). Almost all olivine compositions possess significantly higher NiO and FeO/MnO than expected for olivine in peridotite-derived magmas (Fig. 1b, c and Supplementary Fig. 1), suggesting a contribution of melts from pyroxenitic sources¹⁸. Alternative explanations for these observations seem less plausible (see Methods for discussion). Our interpretation of the olivine compositions implies that the source of the Siberian Traps contained 10–20 wt% recycled oceanic crust (Methods). More specifically, all lavas erupted during the first stage of magmatic activity (Gudchikhinskaya and earlier suites of the Norilsk area) are depleted in heavy rare-earth elements^{19,20}, indicating residual garnet and derivation within or below the base

of thick lithosphere (more than 130 km depth)¹⁴. The source of Gudchikhinskaya lavas was probably almost entirely pyroxenitic¹⁴ (Fig. 1b–d). Younger magmas are not depleted in heavy rare-earth elements, indicating their formation at shallow depths and marked thinning of the lithosphere. Our calculation suggests that these magmas had a near-constant proportion of pyroxenite-derived melt of about 50% (Fig. 1d and Supplementary Table 1) and were strongly contaminated by the continental crust²⁰. Because the main Norilsk section spans less than 1 Myr (ref. 1), it is likely that the lithosphere was thinned in only a few hundred thousand years.

High mantle temperatures over a vast area (Fig. 1a) are consistent with the head of a hot mantle plume^{6,9,17}. On the basis of the petrological constraints we develop a thermomechanical model of the interaction of the plume and lithosphere (see Methods). We assume that the plume arrived below the lithosphere at about 253 Myr ago (model time 0), perhaps near the northern border of the Siberian Shield, where the hottest melts (meimechites) erupted¹⁷. We further assume that the plume head was hot ($T_p = 1,600$ °C; 250 °C excess temperature) and contained a high content (15 wt%) of recycled oceanic crust. In our two-dimensional model we approximate the plume head by a half-circle of radius 400 km located below cratonic lithosphere of variable thickness corresponding to the margin of the Archean craton (130–250 km of depleted lithosphere and 160–250 km of thermal lithosphere; Fig. 2b and Supplementary Fig. 2).

The arrival of a large and hot mantle plume head at the base of the lithosphere has been predicted^{6,21} to cause about 0.8–1 km of broad surface uplift per 100 °C of plume excess temperature. For a purely thermal plume with an excess temperature of 250 °C, we do indeed obtain about 2.0 km of surface uplift (Fig. 2a, red curve). However, if a large fraction (15 wt%) of dense recycled material is present within the plume, its buoyancy is strongly decreased, resulting in little regional uplift (250 m) (Fig. 2a, black curve). Other processes leading to surface subsidence, such as the plume-induced rise of the 670-km phase boundary²² or the crystallization and evacuation of melts, may easily counteract such a small uplift.

The plume head erodes the lowest part of the thermal lithosphere and rapidly spreads below the more refractory depleted lithosphere (Fig. 2b). Its ascent leads to progressive melting of recycled eclogitic material in the plume and to the formation of reaction pyroxenite, which melts at depths of 130–180 km, well before the peridotite (Fig. 2e). The early, purely pyroxenite-derived, melts yielded the lavas of the Gudchikhinskaya and earlier suites that display the ‘garnet signature’ (Fig. 1d).

We propose that massive intrusion of the Gudchikhinskaya suite by dykes imposed compressive stress in the upper brittle part of the lithosphere, ‘locking’ it to the magma transport (Fig. 2b, e shows the moment of ‘locking’). After that, the melt could intrude only into the lower

¹Deutsches GeoForschungsZentrum GFZ, Telegrafenberg, 14473, Potsdam, Germany. ²O.Yu. Schmidt Institute of the Physics of the Earth, Russian Academy of Sciences, 10 ul. B. Gruzinskaya, Moscow, 123995, Russia. ³ISTerre, CNRS, University Joseph Fourier, Maison des Géosciences, 1381 rue de la Piscine, BP 53, 38041 Grenoble Cedex 9, France. ⁴Max Planck Institute for Chemistry, 27 J.-J.-Becher-Weg, Mainz, 55128, Germany. ⁵V. I. Vernadsky Institute of Geochemistry and Analytical Chemistry, Russian Academy of Sciences, 19 ul. Kosygin, Moscow, 119991, Russia. ⁶V. S. Sobolev Institute of Geology and Mineralogy, Siberian Branch of Russian Academy of Sciences, 3 prosp. Akad. Koptyuga, Novosibirsk, 630090, Russia. ⁷Limited Liability Company ‘Norilskgeologiya’ Norilsk, PO Box 889, 663330, Russia.

*These authors contributed equally to this work.

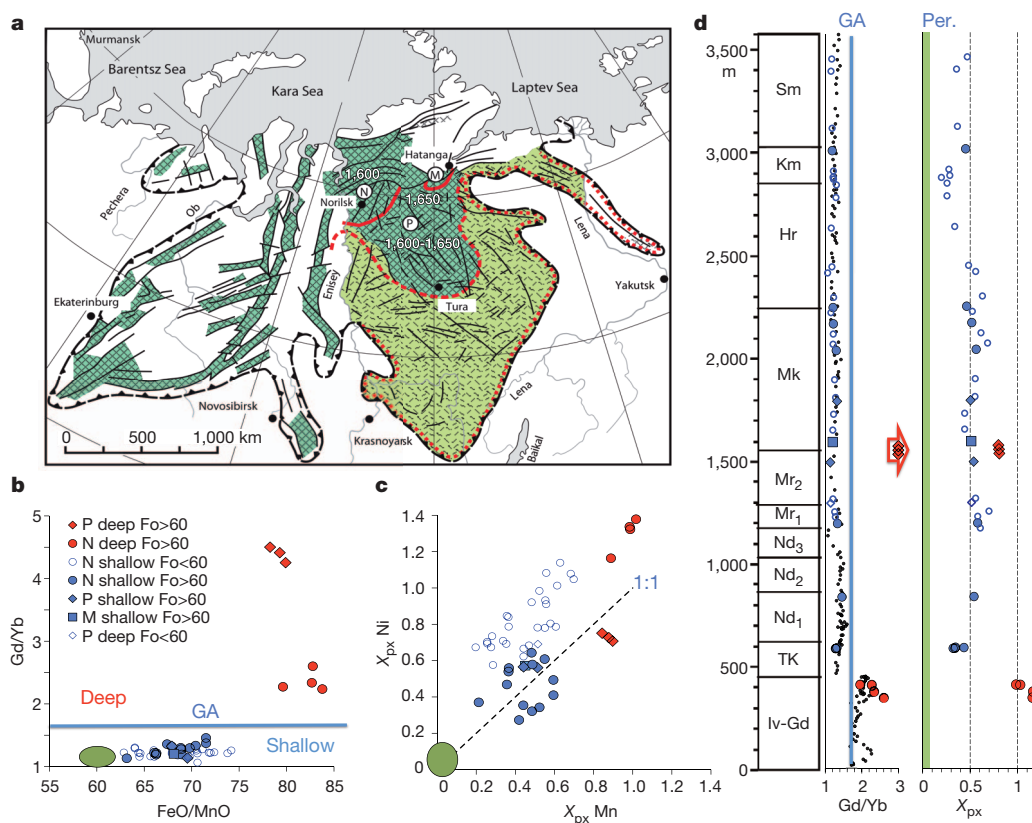


Figure 1 | Petrological constraints. **a**, Geological map of the Siberian Traps³². Dark green areas are lavas, light green areas are tuffs. The dashed black line marks the border of the province. Red lines outline areas with different magmatic activities: solid indicates maximal, dashed is moderate, and dotted is minimal. The three studied regions are Norilsk (N), Putorana plateau (P) and Maymecha-Kotuy province (M). White numbers stand for the potential mantle temperature estimated for lavas of the corresponding areas^{14,17}. **b**, FeO/MnO ratios of olivine phenocrysts over normalized Gd/Yb ratios of host lavas. The blue line marks the pressure that divides 'deep' lavas depleted in heavy rare-earth elements from 'shallow' lavas. The green oval is the reference for the almost pure shallow peridotitic mantle source and indicates the compositions of olivine and lavas from the mid-ocean ridge (Knipovich Ridge, North Atlantic) with minimum amounts of recycled ocean crust in their sources¹⁸. All

olivines are the averages of the three highest Fo percentages of each sample. GA, garnet in the mantle source. **c**, The proportions of pyroxenite-derived melt in the mixture of pyroxenite-derived and peridotite-derived melts calculated independently of Mn deficiency ($X_{px} Mn$) and Ni excess ($X_{px} Ni$) (Methods). **d**, Integrated lava section for Siberian Traps based on the Norilsk section (Supplementary Information). X_{px} is the proportion of pyroxenite-derived melt, calculated as the average of $X_{px} Mn$ and $X_{px} Ni$ for high-forsterite olivines and as $X_{px} Mn$ for low-forsterite olivines, because $X_{px} Ni$ for the latter yields systematic overestimation (Fig. 1c). Small black dots show lavas of the Norilsk section¹⁹. For abbreviations indicating the lava suites of the Norilsk area and normalization for Dy/Yb ratio, see Supplementary Information. Per, peridotite-derived melt component.

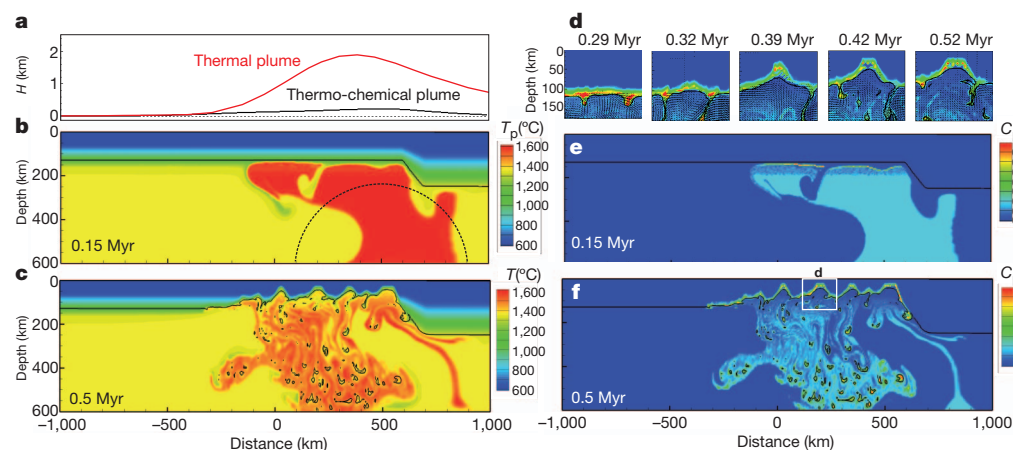


Figure 2 | Model. **a**, Maximum pre-magmatic surface uplift (H) atop a spreading mantle plume with an excess temperature of 250 °C. The red curve corresponds to the purely thermal plume, and the black curve corresponds to a thermo-chemical plume containing 15 wt% of recycled crust. **b**, **c**, Temperature distributions (°C) in the model cross-section at model times of 0.15 Myr (**b**) and 0.5 Myr (**c**). The solid line marks the boundary of the depleted lithosphere, and the dashed half-circle denotes the initial shape of the starting plume.

d, Snapshots of the plume breaking through the lithosphere in the domain shown by the white rectangle in **f**. Colours show concentrations of the pyroxenitic component in the plume or in the crystallized melt. **e**, **f**, Distribution of the pyroxenite component in the plume (C_{px}) or in the crystallized melt in the model cross-section at model times of 0.15 Myr (**e**) and 0.5 Myr (**f**). The solid line marks the boundary of the depleted lithosphere.

lithosphere (see Methods). The intruding melt cools and crystallizes to dense eclogite. It also strongly heats and weakens the lithosphere, promoting Raleigh–Taylor instabilities⁸. The lower part founders, and the base of the lithosphere is mechanically eroded (Fig. 2c). Enriched in eclogite, the lithospheric material in the boundary layer above the plume escapes to the sides of the plume and then downwards, allowing the plume to ascend (Fig. 2d, f). The plume breaks through the lithosphere in several zones, and in only 100–200 kyr reaches its minimum depth of about 50 km (Fig. 2d and Supplementary Fig. 2). At this level, mafic melts crystallize to a garnet-free assemblage and have a density lower than that of the ambient mantle, thus preventing the formation of Raleigh–Taylor instabilities. This mode of rapid lithosphere destruction does not require regional stretching and matches observations for the Siberian Traps^{7,12}.

The extent of lithospheric destruction depends, among other factors (Supplementary Information), on the initial density of mantle lithosphere, which is controlled by its composition (Fig. 3a and Supplementary Information). In the case of re-fertilized or moderately depleted mantle lithosphere, the volume of the melt that intrudes into the crust (melt crossing 50 km depth) reaches few per cent of the plume volume (Fig. 3a), which leads to substantial melting of the crust and contamination of basalts. Using the proportion of the magma-to-plume volumes from a two-dimensional model for a three-dimensional plume head with a radius of 400 km, we estimate the volume of the magma intruded into the crust to be $(6\text{--}8) \times 10^6 \text{ km}^3$, which is realistic for the Siberian Traps^{1,12}.

In agreement with geochemical data, the model predicts that most magma contains about 50% of pyroxenite-derived melt (Fig. 3b, blue curves and symbols) and lack the ‘garnet signature’ because they are generated at depths of less than 60 km. For the melt generated deeper than 100 km, the model predicts a much higher proportion of pyroxenite melt (75–100%; Fig. 3b, red curves), again in agreement with observations (Fig. 3b, red symbols).

Our model allows us to estimate the volume of CO₂ and HCl gases released from the plume. For these calculations we consider separately the recycled crust and peridotitic components by using data from melt inclusions in olivine in Gudchikhinskaya picrites and mantle peridotite as well as published estimates (Methods). For the composition of recycled crust this yields HCl = 137 p.p.m., S = 135 p.p.m., H₂O = 800 p.p.m. and CO₂ > 900 p.p.m. The model predicts that most of the CO₂ and HCl in the recycled-crust component of the plume is extracted during

its interaction with the lithosphere (Fig. 4a), and a major part is extracted before the main phase of magmatism. For a three-dimensional plume with a radius of 400 km, the mass of extracted CO₂, which comes mostly from the recycled component of the plume, is more than 170×10^{12} tonnes. This is several times larger than previous estimates^{13,15} and also exceeds the maximum estimate of the CO₂ released from the magmatic heating of the coals from the Tunguska basin¹⁵.

Our prediction of the mass of CO₂ extracted from the plume is consistent with the amount of CO₂ released during the Permo-Triassic mass extinction estimated from Ca isotope data²³ (Fig. 4a). Moreover, if we use a $\delta^{13}\text{C}$ value of -12‰ for pyroxenite-derived melt as measured in Koolau (Hawaii) basaltic melt for the source dominated by the recycled crust component²⁴, we can also explain the ¹³C excursion associated with the main mass-extinction event²³ (Methods). Therefore CO₂ from the plume alone may have triggered the main extinction event. We speculate that low-density and low-viscosity volatiles were the first to penetrate the compressed and mechanically locked crust, triggering the extinction (Fig. 4a, upper axis). Alternatively, a sufficient quantity of gases may have been released before lithospheric locking, together with the deep-sited magmas of pre-Gudchikhinskaya suites, which could also produce metamorphic gases by magmatic heating of the coals and carbonates. This alternative is supported by the recent discovery of coal fly ash in Permian rocks from the Canadian High Arctic immediately before the mass extinction, interpreted as a result of combustion of Siberian coal and organic-rich sediments by flood basalts²⁵. In either case, according to our model, the major mass extinction happened before the main phase of flood basalt extrusion. In contrast, most of the CO₂ and other gases released by contact metamorphism of carbon-rich and sulphur-rich sediments, which have been suggested as a trigger for the mass extinction^{15,16}, would be released during the main phase of magmatism. Precise U–Pb dating of Siberian magmatic units and the Permo-Triassic boundary is required to distinguish between the two hypotheses. Nonetheless, existing geochronological data^{26,27} and the presence of abundant pyroclastic rocks underlying lavas of the main magmatic phase^{11,19} support the idea that the major mass extinction predated the main phase of magmatism (see Fig. 4a, upper axis). Additional large amounts of gases released from heated sediments^{15,16} may have been the cause of ¹³C excursions during the later phases of the biotic crisis²⁸.

According to our data and model, the plume also generates a surprisingly large amount of HCl (about 18×10^{12} tonnes; Fig. 4a), mostly

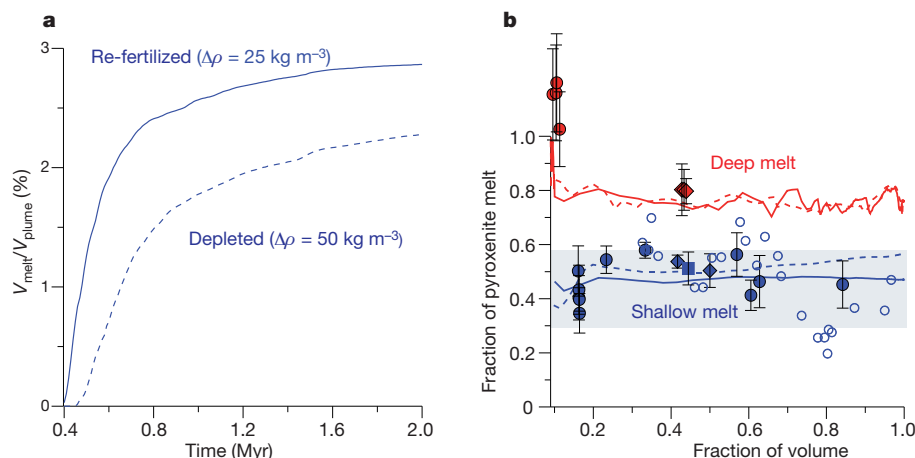


Figure 3 | Model predictions. **a**, Evolution in time of a melt volume crossing the 50-km depth and normalized to the volume of the plume. The solid and dashed curves correspond to the models with re-fertilized lithosphere and moderately depleted lithosphere, respectively. **b**, Plot of the fraction of pyroxenitic component in basalts of the Norilsk cross-section against the fraction of the volume of extruded magmas. The blue colour corresponds to the ‘shallow’ melts that do not retain a garnet signature; the red colour corresponds to the deep melts that retain a garnet signature. Symbols denote data from

olivine compositions; see Fig. 1b for details. Error bars correspond to 1 standard deviation of the mean of pyroxenite-derived melt proportions estimated independently from Ni excess and Mn deficiency of olivine (Methods and Supplementary Table 1). The solid and dashed curves show the modelled average melt compositions with re-fertilized and moderately depleted lithosphere, respectively. The grey rectangle shows the range of variation of the melt compositions predicted by the model.

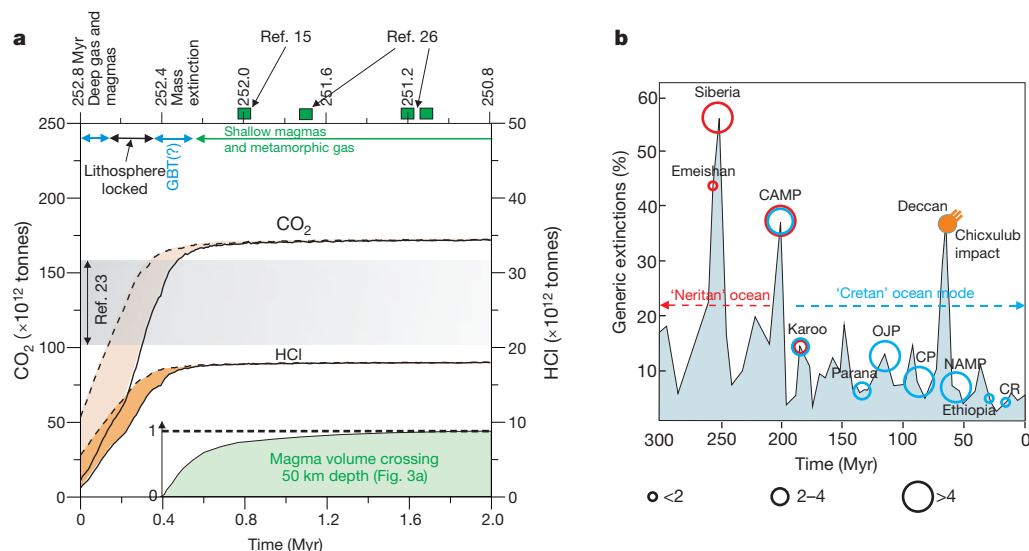


Figure 4 | Production of volatiles and its consequences for mass extinctions.

a, Plot of modelled CO_2 (left axis) and HCl (right axis) amounts extracted from the plume against model time (lower axis). Solid curves show the minimum estimate and dashed curves the maximum estimate of CO_2 and HCl extracted from the plume (Methods). The grey rectangle shows the estimated range of the released CO_2 during the Permo-Triassic mass extinction²³. The green area shows time dependence of the normalized volume of the magma crossing the 50-km depth, calculated for the re-fertilized lithosphere (Fig. 3a). On the top axis we show geological time and a possible model for triggering the Permo-Triassic mass

extinction. GBT, gases break through. Also shown is U–Pb dating of the extinction event²⁷ and U–Pb dating of main-phase Siberian basalts²⁶ and intrusions¹⁵. **b**, Plot of mass extinction intensity (light blue field) with major LIPs (circles) against geological time (modified from ref. 33), together with the timing of different ocean modes³⁰. Circle colours denote the timing of LIPs relative to ocean modes: blue, ‘Cretan’ mode; red ‘Neritan’ mode; blue and red together, transition mode. The scale of circle sizes is in millions of cubic kilometres. CAMP, Central Atlantic Magmatic Province; NAMP, Northern Atlantic Magmatic Provinces, OJP, Ontong Java; CP, Caribbean Plateaux; CR, Columbian River basalts.

also derived from the recycled component. This quantity of toxic HCl must have been extremely damaging for the terrestrial species and was also sufficient to trigger deadly instability of the stratospheric ozone layer²⁹. By accepting our viewpoint that degassing of the plume, rather than thermogenic gases from sediments, triggered the biotic crises, we lose an elegant explanation of why the Siberian LIP was so much more damaging to biota than other LIPs of comparable size (Karoo, Parana and North Atlantic) that extruded through other types of sediment or granitic rock^{15,16}. An alternative explanation is based on the correlation of the intensity of mass extinctions with the age of Phanerozoic LIPs (Fig. 4b), a relationship that can be explained by the temporally different response of the ocean to acidification by the large amounts of released CO_2 (refs 23, 30). In contrast to the pre-Mid-Mesozoic ‘Neritan’ ocean, the more recent ‘Cretan’ ocean was buffered against acidification by deep-sea unlithified carbonate sediments and was thus much more resistant to acidification^{23,30}. Therefore, CO_2 degassing of a pre-Mid-Mesozoic LIP caused much more severe ocean acidification and mass extinction than later LIPs (Fig. 4b). The only exception is the Deccan LIP and the contemporaneous mass extinction at 65.5 Myr ago; however, in this case the Chicxulub impact was an additional contributing factor³¹.

Numerical tests (Supplementary Figs 4–6) suggest that rapid lithospheric destruction associated with melting in the heads of thermochemical plumes is valid for the large range of plume parameters and lithospheric thicknesses, and therefore may apply not only to the Siberian Traps but also to other LIPs. An absence of prominent pre-magmatic uplift does not argue against a plume origin of LIPs, but may instead point to a high content of recycled crust within the plume. In such cases, other parameters being equal, the model predicts that eclogite-rich plumes caused the most extensive delamination and thinning of the lithosphere, thus best preparing it for a possible break-up. They also produced the strongest volcanism and led to the most marked climatic consequences.

Another suggestion of our model—that major mass extinctions are triggered by degassing of plume magmas that predate the main magmatic phase—also seems to be consistent with the observations for many LIPs⁵,

implying that gas output from plume heads may be much larger than previously thought.

METHODS SUMMARY

We report new data on 45 representative olivine-bearing samples of Siberian flood basalts from the Norilsk, Putorana and Maimecha–Kotui regions (Supplementary Tables 1 and 3). The bulk rocks were crushed, melted and analysed for major and trace elements with an electron probe microanalyser (EPMA) and by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) at the Max Planck Institute for Chemistry in Mainz, Germany. Olivine phenocrysts (about 2,500 analyses) were analysed by EPMA with a special high-precision protocol¹⁸ at the Max Planck Institute for Chemistry. Using this new information and published approaches we estimated the amount of recycled oceanic crust in the sources of basalts and their potential temperatures, and discuss possible alternative models of the source compositions. We further estimate the amounts of H_2O , Cl , S and CO_2 in the recycled oceanic crust and develop a model of its degassing during plume–lithosphere interaction.

We model the thermomechanical interaction of the plume and lithosphere by numerically solving a coupled system of momentum, mass and energy conservation equations in two dimensions. We employ nonlinear temperature and stress-dependent elasto-visco-plastic rheology, consider pressure-dependent and temperature-dependent melting of a heterogeneous mantle and employ simple models of melt transfer and extraction of volatiles.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 February; accepted 26 July 2011.

- Reichow, M. K. *et al.* The timing and extent of the eruption of the Siberian Traps large igneous province: implications for the end-Permian environmental crisis. *Earth Planet. Sci. Lett.* **277**, 9–20 (2009).
- White, R. & McKenzie, D. Magmatism at rift zones—the generation of volcanic continental margins and flood basalts. *J. Geophys. Res. Solid Earth Planets* **94**, 7685–7729 (1989).
- Garfunkel, Z. Formation of continental flood volcanism—the perspective of setting of melting. *Lithos* **100**, 49–65 (2008).
- Courtillot, V. E. & Renne, P. R. On the ages of flood basalt events. *C. R. Geosci.* **335**, 113–140 (2003).
- Wignall, P. B. Large igneous provinces and mass extinctions. *Earth Sci. Rev.* **53**, 1–33 (2001).
- Campbell, I. H. & Griffiths, R. W. Implications of mantle plume structure for the evolution of flood basalts. *Earth Planet. Sci. Lett.* **99**, 79–93 (1990).

7. Czamanske, G. K., Gurevitch, A. B., Fedorenko, V. & Simonov, O. Demise of the Siberian plume: paleogeographic and paleotectonic reconstruction from the prevolcanic and volcanic record, north-central Siberia. *Int. Geol. Rev.* **40**, 95–115 (1998).
8. Elkins-Tanton, L. T. & Hager, B. H. Melt intrusion as a trigger for lithospheric foundering and the eruption of the Siberian flood basalts. *Geophys. Res. Lett.* **27**, 3937–3940 (2000).
9. Richards, M. A., Duncan, R. A. & Courtillot, V. E. Flood basalts and hot-spot tracks—plume heads and tails. *Science* **246**, 103–107 (1989).
10. Cordery, M. J., Davies, G. F. & Campbell, I. H. Genesis of flood basalts from eclogite-bearing mantle plumes. *J. Geophys. Res. Solid Earth* **102**, 20179–20197 (1997).
11. Sobolev, V. S. *Petrology of Siberian Traps* (Transactions of All-Union Arctic Institute, vol. 43) (Glavnogo Upravleniya Sevmorputi, 1936).
12. Dobretsov, N. L., Kiryashkin, A. A., Kiryashkin, A. G., Vernikovskiy, V. A. & Gladkov, I. N. Modelling of thermochemical plumes and implications for the origin of the Siberian traps. *Lithos* **100**, 66–92 (2008).
13. Self, S., Widdowson, M., Thordarson, T. & Jay, A. E. Volatile fluxes during flood basalt eruptions and potential effects on the global environment: a Deccan perspective. *Earth Planet. Sci. Lett.* **248**, 518–532 (2006).
14. Sobolev, A. V., Krivolutsкая, N. A. & Kuzmin, D. V. Petrology of the parental melts and mantle sources of Siberian trap magmatism. *Petrology* **17**, 253–286 (2009).
15. Svensen, H. *et al.* Siberian gas venting and the end-Permian environmental crisis. *Earth Planet. Sci. Lett.* **277**, 490–500 (2009).
16. Ganino, C. & Arndt, N. T. Climate changes caused by degassing of sediments during the emplacement of large igneous provinces. *Geology* **37**, 323–326 (2009).
17. Sobolev, A. V., Sobolev, S. V., Kuzmin, D. V., Malitch, K. N. & Petrunin, A. G. Siberian meimechites: origin and relation to flood basalts and kimberlites. *Russ. Geol. Geophys.* **50**, 999–1033 (2009).
18. Sobolev, A. V. *et al.* The amount of recycled crust in sources of mantle-derived melts. *Science* **316**, 412–417 (2007).
19. Fedorenko, V. A. *et al.* Petrogenesis of the flood-basalt sequence at Noril'sk, North Central Siberia. *Int. Geol. Rev.* **38**, 99–135 (1996).
20. Wooden, J. L. *et al.* Isotopic and trace-element constraints on mantle and crustal contributions to Siberian continental flood basalts, Norilsk Area, Siberia. *Geochim. Cosmochim. Acta* **57**, 3677–3704 (1993).
21. Farnetani, C. G. & Richards, M. A. Numerical investigations of the mantle plume initiation model for flood-basalt events. *J. Geophys. Res. Solid Earth* **99**, 13813–13833 (1994).
22. Leng, W. & Zhong, S. J. Surface subsidence caused by mantle plumes and volcanic loading in large igneous provinces. *Earth Planet. Sci. Lett.* **291**, 207–214 (2010).
23. Payne, J. L. *et al.* Calcium isotope constraints on the end-Permian mass extinction. *Proc. Natl Acad. Sci. USA* **107**, 8543–8548 (2010).
24. Hauri, E. SIMS analysis of volatiles in silicate glasses. 2. Isotopes and abundances in Hawaiian melt inclusions. *Chem. Geol.* **183**, 115–141 (2002).
25. Grasby, S. E., Sanei, H. & Beauchamp, B. Catastrophic dispersion of coal fly ash into oceans during the latest Permian extinction. *Nature Geosci.* **4**, 104–107 (2011).
26. Kamo, S. L. *et al.* Rapid eruption of Siberian flood-volcanic rocks and evidence for coincidence with the Permian–Triassic boundary and mass extinction at 251 Ma. *Earth Planet. Sci. Lett.* **214**, 75–91 (2003).
27. Mundil, R., Palfy, J., Renne, P. R. & Black, P. The Triassic timescale: new constraints and a review of geochronological data. *Geol. Soc. Lond. Spec. Pub.* **334**, 41–60 (2010).
28. Payne, J. L. & Kump, L. R. Evidence for recurrent Early Triassic massive volcanism from quantitative interpretation of carbon isotope fluctuations. *Earth Planet. Sci. Lett.* **256**, 264–277 (2007).
29. Beerling, D. J., Harfoot, M., Lomax, B. & Pyle, J. A. The stability of the stratospheric ozone layer during the end-Permian eruption of the Siberian Traps. *Phil. Trans. R. Soc. Lond. A* **365**, 1843–1866 (2007).
30. Ridgwell, A. A mid Mesozoic revolution in the regulation of ocean chemistry. *Mar. Geol.* **217**, 339–357 (2005).
31. Schulte, P. *et al.* The Chicxulub asteroid impact and mass extinction at the Cretaceous–Paleogene Boundary. *Science* **327**, 1214–1218 (2010).
32. Masaitis, V. L. Permian and Triassic volcanism of Siberia: problems of dynamic reconstructions [in Russian]. *Zapiski Vsesoyuznogo Mineralogicheskogo Obshchestva* **112**, 412–425 (1983).
33. White, R. V. & Saunders, A. D. Volcanism, impact and mass extinctions: incredible or credible coincidences? *Lithos* **79**, 299–316 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements S.V.S. and A.V.S. are especially grateful to Vladimir Stepanovich Sobolev, who excited their interest in the origin of the Siberian Traps. We thank G. A. Fedorenko for providing data on the Norilsk lavas and for discussions; N. Groschopf for help in managing the electron probe microanalyser; O. Kuzmina, N. Svirskaya and T. Shlichikova for sample preparation; P. Cardin, N. Dobretsov, E. Galimov, C. Herzberg, A. Hofmann, L. Kogarko, H.-C. Nataf, J. Payne, Y. Podladchikov, I. Ryabchikov, A. Turchyn and G. Wörner for discussions; and P. Kelemen for comments. S.V.S. thanks the Deutsche Forschungsgemeinschaft (DFG) SPP 1375 SAMPLE (SO 425/4) for support. The study by A.V.S. was funded by the Agence Nationale de la Recherche, France (Chair of Excellence Grant ANR-09-CEXC-003-01) and partly supported by a Gauss Professorship in Göttingen University, Germany, the Russian Foundation for Basic Research (09-05-01193a), a Russian President grant for leading Russian scientific schools (HIII-3919.2010.5) and an Earth Sciences Department of Russian Academy Grants.

Author Contributions S.V.S. and A.V.S. provided major contributions to thermomechanical (S.V.S.) and petrological (A.V.S.) modelling, to the interpretation of data and to the writing of the paper. N.A.K. provided geological background and contributed to interpretation. A.G.P. contributed to the thermomechanical modelling at an initial stage. N.T.A. contributed to interpretation and writing of the paper. D.V.K. processed samples and performed the measurements. N.A.K., V.A.R. and Y.R.V. provided carefully selected samples. All authors contributed intellectually to the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.V.S. (stephan@gfz-potsdam.de) or A.V.S. (alexander.sobolev@ujf-grenoble.fr).

METHODS

Samples. The samples studied and their localities are described in Supplementary Information.

Analytical methods. Olivine grains were manually separated from crushed lavas, then mounted and polished in epoxy. The compositions of olivine were analysed with an EPMA on a Jeol JXA 8200 SuperProbe at the Max Planck Institute for Chemistry (Mainz, Germany) at an accelerating voltage of 20 kV and a beam current of 300 nA, following a special procedure which allows 20–30 p.p.m. (2 σ error) precision and accuracy for Ni, Ca, Mn, Al, Ti, Cr and Co, and 0.02 mol% accuracy for the forsterite component in olivine¹⁸.

Bulk rocks were crushed, melted to glass³⁴ and then mounted and polished in epoxy. Major and trace elements were also determined by EPMA at the Max Planck Institute for Chemistry. Major-element abundances in glasses were measured at an accelerating voltage of 15 kV and a beam current of 12 nA with a reference sample of natural basaltic glass USNM111240/52 (VG2)³⁵, with a relative error of 1–2%. LA-ICP MS was used to determine trace elements in glasses of melt inclusions and in olivines, on an ELEMENT-2, Thermo Scientific mass spectrometer with a UP-213 New Wave Research solid-phase laser at the Max Planck Institute for Chemistry, with reference to the KL-2G and NIST 612 standard samples of basaltic glass³⁶.

Proportions of pyroxenite-derived melt and recycled crust. We interpret excessive Ni and deficient Mn concentrations in Siberian olivine phenocrysts relative to olivine in peridotite-derived melt as a result of the contribution of olivine free pyroxenite lithology in their source^{18,37–39}. The alternative explanations of this phenomenon are discussed briefly in the next section. Relative proportions of pyroxenite and peridotite derived melts were estimated from MnO/FeO (X_{px} , Mn) and NiO \times FeO/MgO (X_{px} , Ni) ratios³⁸ for each sample, using the average composition of most magnesian olivines (defined by olivines with Fo within 3 mol% from a maximum Fo).

The amount of recycled crust in the plume is linked to the proportion of pyroxenite-derived melt by the degree of melting of the eclogite component, the amount of eclogite-derived melt needed to produce hybrid pyroxenite from peridotite, and the degrees of melting of peridotite and pyroxenite^{18,39}. We calculated the amount of recycled crust in the Siberian plume using the approach described previously¹⁸ and their equation S3, and the following assumptions: a maximum degree of melting of eclogite and pyroxenite of 60%; an average proportion of pyroxenite-derived melt in shallow magmas of 46% (Supplementary Table 1); and melting of peridotite at 50 km depth. The amounts of recycled crust are 10% and 20% for 10% and 25% melting of peridotite, respectively.

Alternative explanations of the unusual olivine composition. Alternative explanations of high Ni/Mg and FeO/MnO ratios in olivine include: (1) effect of clinopyroxene crystallization, (2) an underestimated temperature effect on olivine-melt partition of Ni, and (3) contribution of core material to the mantle source. None of these alternatives require a significant role of olivine-free pyroxenite in the mantle source. In addition there are different models for the pyroxenite origin, which may affect our estimation of proportions of pyroxenite in the Siberian plume. These include (4) solid-state reaction of peridotite and recycled crust in the lower mantle and (5) partial reaction between eclogite-derived melt and peridotite. Below we briefly discuss these alternatives.

1. Crystallization of clinopyroxene together with olivine may increase both Ni/Mg and Fe/Mn ratios in olivine. This effect could be particularly important for the low magnesian evolved olivine. However clinopyroxene crystallization can be recognized by low Ca in olivine. In Supplementary Information we show that this is an unlikely situation for most studied olivines, which do not show a significant decrease in Ca concentrations. In particular, early clinopyroxene crystallization cannot explain the composition of the olivines from picrites of the Gudchikhinskaya formation and the Ayan river, which are extremely rich in Ni and deficient in Mn. In addition, melt inclusions in the former¹⁴ and olivines in the latter (Supplementary Table 3) do not indicate early clinopyroxene crystallization. We conclude that the fractionation of clinopyroxene could not have produced the observed anomalies in the olivine compositions.

2. An underestimated temperature effect on olivine-melt partition coefficient, if present, may increase Ni concentration in the shallow olivine compared with olivine in the deep source as a result of temperature difference between the sites of generation and crystallization. This issue has been discussed previously^{37,39}, where it was shown that any temperature effect additional to the compositional one considered in the model of Ni partitioning between melt and olivine used in this study⁴⁰ is too small to explain the extent of Ni excess observed in Siberian olivines.

3. The contribution of core material to increase Ni/Mg (ref. 41) and Fe/Mn (ref. 42) ratios has been discussed previously¹⁸, where it was shown that this explanation is highly unlikely because of a lack of correlation between Ni excess and high Co concentrations in the olivines.

4. The solid-state reaction between recycled crust and peridotite in the lower mantle may produce pyroxenite with a composition much closer to that of peridotite, which in the upper mantle will be transformed to olivine-bearing pyroxenite³⁷. If this lithology exists, it could potentially be the source of parental melts of typical Siberian basalts. However, olivine-bearing lithologies could not be the source of deep-sourced lavas such as the Gudchikhinskaya formation and Ayan river picrites, whose olivine compositions demonstrate derivation from a dominating olivine-free source. In addition, solid-state reactions of the type envisaged in ref. 37 will be limited by slow volume diffusion in the mantle⁴³, whereas the production of reaction olivine-free pyroxenite will be restricted only by melt percolation velocity^{18,39}.

5. Incomplete reaction between eclogite-derived melt and peridotite may produce olivine-bearing lithologies⁴⁴, which could be potential sources of the parental melts of typical Siberian traps. However, these olivine-bearing lithologies cannot produce Siberian deep-sourced magmas (see above).

We conclude that although we cannot fully exclude some of proposed alternatives, our explanation of the olivine compositions is based on solid grounds and seems the most plausible.

Potential temperatures. The published potential temperature of 1,540 °C for the source of Gudchikhinskaya magmas¹⁴ has been corrected for the effect of 40% melting of pyroxenitic source¹⁷, the amount of pyroxenite in the plume (15%) and the latent heat of melting⁴⁵. The value obtained is about 1,600 °C. Potential temperatures for the Maimecha–Kotuy province¹⁷ and the Putorana plateau⁴⁶ (see Fig. 1a) were obtained for magmas strongly enriched in highly incompatible elements. These magmas originated at low degrees of melting and thus were not corrected for the melting effect.

Amount of volatile elements. The concentrations of volatile elements in the recycled oceanic crust in the Siberian plume were constrained using the compositions of inclusions of uncontaminated melt in early olivine phenocrysts from Gudchikhinskaya magmas. For these magmas it was shown from both olivine and melt compositions that they probably represent the melting of a pure pyroxenitic source¹⁴. Inclusions in olivine from these magmas have been shown to represent primary melts¹⁴ and thus their concentrations of Cl, S and H₂O can be used to estimate the contents of these volatiles in the mantle source. For the composition of source eclogite, this yields the following values^{14,17}: Cl = 137 p.p.m., S = 135 p.p.m. and H₂O = 800 p.p.m., after normalization of the values to K. In the deep mantle, these amounts of Cl, S and H₂O could reside in chloride⁴⁷, sulphide, and garnet or pyroxene respectively. In contrast with Cl, S and H₂O, the amount of CO₂ in relatively shallow melts does not represent the primary concentration, because of almost complete degassing at high pressures. Thus for an assessment of CO₂ in the recycled oceanic crust we use global estimations of 3,000 p.p.m. CO₂ for the bulk 7-km-thick oceanic crust and its maximum outgassing rate through arc volcanism of 70% (ref. 48). This gives a conservative minimum estimate of 900 p.p.m. CO₂ in the deeply recycled oceanic crust. The maximum estimate would be about 1,800 p.p.m. using the same initial bulk concentrations of CO₂ and minimal outgassing of 40% (ref. 48). In the deep mantle this amount of CO₂ could reside in carbonates or diamond⁴⁷. For our model we use the minimum conservative estimate of CO₂ = 900 p.p.m.

Thermo-mechanical numerical technique. We use a fully coupled thermo-mechanical formulation for the system of momentum, mass and energy conservation equations in two dimensions with nonlinear temperature and stress-dependent elasto-visco-plastic rheology, described in detail in ref. 49 (for parameters and procedure for calculating density, see Supplementary Information). Equations are solved numerically using the explicit Lagrangean FEM technique LAPEX2D (ref. 50) based on a FLAC algorithm (prototype described in ref. 51) combined with a particle-in-cell approach. All time-dependent fields including full stress tensor are stored at particles.

Melting models. We use a simplified model for batch melting of four components: peridotite, pyroxenite and two eclogites formed through the crystallization of peridotitic and pyroxenitic melts, respectively. Melting temperatures are defined as follows: for peridotite we use the dry batch melting model⁵²; for pyroxenite we use the following relation from experiments¹⁸ for dry batch melting of pyroxenite:

$$T_{\text{m}}^{\text{px}} = 976 + 12.3P - 0.051P^2 + 663.8X - 611.4X^2$$

where P is pressure in kbars, T is potential temperature in °C and X is degree of melting, $0 < X < 0.55$.

For eclogite of both types (peridotite-derived or pyroxenite-derived) we use the following relations approximating experiments⁵³ for dry batch melting of eclogite at 50% melting:

$$T_{\text{m}}^{\text{c}} = 1173.4 + 5.78P$$

at $P < 55$ kbar, and

$$T_{\text{m}}^{\text{c}} = -237.5 + 48.0P - 0.3P^2$$

at $55 \leq P < 80$ kbar.

The melting for each finite element is organized sequentially, beginning with the component with the lowest melting temperature (usually eclogite), then usually pyroxenite and finally peridotite. If the current temperature (T) in a finite element exceeds the melting temperature (T_m) of the component that exists in this element, then a certain amount of melt (dC_m) is generated that lowers the current temperature to the melting temperature, $dC_m = (T - T_m)C_p/(\Delta S_m T)$, where C_p is heat capacity and ΔS_m is entropy of melting, set to $1,200 \text{ J kg}^{-1} \text{ K}^{-1}$ and $400 \text{ J kg}^{-1} \text{ K}^{-1}$, respectively, for all components.

Model for melt transfer. We assume that melt transfer within the melting domain occurs much faster than the Raleigh–Taylor instability develops in the lithosphere. This assumption holds if the velocity of melt transfer (V_m) is much higher than H/τ , where H is the typical distance of melt transfer and τ is typical time of Raleigh–Taylor instability. With values of $H < 50 \text{ km}$ and $\tau \approx 50,000$ years (see Supplementary Fig. 2), this assumption is valid if V_m is much higher than 1 m yr^{-1} . In reality, V_m is at least tenfold higher in the upper mantle regions where intensive melting occurs⁵⁴.

If present, the entire melt is assumed to move rapidly upwards within the domain where the local temperature is higher than the melting temperature. As usual for melt porous flow, we also assume that it is thermally equilibrating at each element. In practice, at every n th calculation time-step, for every element we check the melting condition and move the entire volume of melt (if present) one element upwards, recalculating the temperature of that element according to the local energy conservation law.

For the melt in the uppermost elements of the melting domain, we consider two transfer modes. First we consider a mode that mimics transfer through fractures: in this case, a fraction or the entire melt is assumed to move to the surface; in practice it is just taken out from the model. Second, we consider a mode of mechanically locked lithosphere: in this case the entire melt from each uppermost element of the melting zone is moved to, and evenly distributed between, K elements in the column just above it (usually we take $K = 4$) and is assumed to crystallize. Simultaneously, the rheology of melt-accepting lithospheric elements is switched from ‘dry’ to ‘wet’ olivine rheology if the crystallized melt content exceeds some critical value (we take this value as 1%). The temperature and composition of the accepting elements are recalculated according to the local energy and mass conservation laws.

Extraction of volatiles. We consider two endmember models for the extraction of volatiles. In the first model we assume that CO_2 and HCl are fully extracted from the plume if the temperature approaches that of the carbonatite solidus⁵⁵. This model gives an upper bound for melt mobility, assuming that melts produced by an infinitely low degree of melting can move out of the plume. In the second model we assume that CO_2 and HCl are fully extracted from both peridotitic and pyroxenitic components only if 1% melting is achieved. This model gives lower boundary for melt mobility, assuming that only 1% carbonate–silicate melts can move out of the plume. For both models we assume a concentration of HCl in recycled crust of 137 p.p.m. derived from melt inclusions in olivine, no HCl in the peridotitic component and minimum conservative estimates of CO_2 content in both recycled crust (900 p.p.m.) and plume peridotite (70 p.p.m.) (see above).

Expected mode of motion of volatiles in the lithosphere. Melt from the plume is trapped and crystallizes in the lithosphere; it then returns to the mantle as the lithosphere founders. The volatiles extracted by melting of the plume are released as the melt crystallizes because host phases are not stable at the high temperatures at the base of the lithosphere. They migrate upwards, then react and are fixed in carbonates or chlorides in the cooler upper part of the lithosphere. Continuous upward migration of high-temperature isotherms then decomposes these phases, promoting further displacement of the volatile front ahead of the basalt-melting front. According to our model, more than 70% of the mafic magmas generated in the plume crystallized to eclogites and subsided back into the mantle as the densified lithosphere foundered; less than 30% were intruded into the crust. However, almost all carbonatite melts traversed the lithosphere and crust, because the temperature of the detached blocks was significantly higher than the melting temperature of carbonatite. It is therefore likely that a significant part of the volatiles released from the plume finally reached the surface, promoting explosive eruptions, which are very common in the early stage of Siberian Traps^{11,19} and other LIPs, that is, Emeishan flood basalts⁵⁶. Note that the volatiles that were initially stored in the minerals of the destructed portions of the lithosphere should be also melted out and could finally reach the surface as well.

Why output of volatiles from LIPs could be drastically underestimated. The amount of volatiles released was previously estimated using only the volume of extruded magmas or magmas intruded into the shallow crust. In these studies were disregarded the magmas that crystallized in the deep crust and, more importantly,

the much larger volumes of magmas that we propose were involved in the destruction of the lithosphere and never reached the crust, although most of the volatiles extracted from them probably did (see above). Additionally, the recycled crust component of the plume, which contains much more volatiles than the peridotitic component, was not previously considered in the balance calculation.

Estimation of ^{13}C excursion. The ^{13}C -isotope change due to the released CO_2 (considered to be instantaneous) can be estimated from a simple mass-balance equation³⁸. According to our model, about 172×10^{12} tonnes of CO_2 is released from the plume—about 70% from recycled crust and the remaining 30% from peridotite. Assuming that $\delta^{13}\text{C} = -12\text{‰}$ for crust-derived CO_2 (ref. 24) and $\delta^{13}\text{C} = -5\text{‰}$ for peridotite-derived CO_2 , we obtain an average isotopic composition of the plume-released carbon of $\delta^{13}\text{C} = -9.9\text{‰}$. Assuming $\delta^{13}\text{C} = 3.6\text{‰}$ for the initial carbon isotope composition and 300×10^{12} tonnes of CO_2 (or 82×10^{12} tonnes of C)²³ for the Late Permian CO_2 reservoir, we estimate that the magnitude of the carbon isotope excursion was 4.9‰ if all plume-released gases migrated to the surface, and 3.5‰ if only half of them arrived. Both numbers are well within the range of reported values for the Permian–Triassic excursion⁵⁷.

34. Stoll, B. *et al.* An automated iridium-strip heater for LA-ICP-MS bulk analysis of geological samples. *Geostand. Geoanal. Res.* **32**, 5–26 (2008).
35. Jarosevich, E. J., Nelen, J. A. & Norberg, J. A. Reference sample for electron microprobe analysis. *Geostand. Newsl.* **4**, 43–47 (1980).
36. Jochum, K. P. *et al.* The preparation and preliminary characterisation of eight geological MPI-DING reference glasses for in-situ microanalysis. *Geostand. Newsl.* **24**, 87–133 (2000).
37. Herzberg, C. Identification of source lithology in the Hawaiian and Canary Islands: implications for origins. *J. Petrol.* **52**, 113–146 (2011).
38. Sobolev, A. V., Hofmann, A. W., Brugmann, G., Batanova, V. G. & Kuzmin, D. V. A quantitative link between recycling and osmium isotopes. *Science* **321**, 536 (2008).
39. Sobolev, A. V., Hofmann, A. W., Sobolev, S. V. & Nikogosian, I. K. An olivine-free mantle source of Hawaiian shield basalts. *Nature* **434**, 590–597 (2005).
40. Beattie, P., Ford, C. & Russell, D. Partition coefficients for olivine-melt and orthopyroxene-melt systems. *Contrib. Mineral. Petrol.* **109**, 212–224 (1991).
41. Ryabchikov, I. D. High NiO content in mantle-derived magmas as evidence for material transfer from the Earth's core. *Dokl. Earth Sci.* **389**, 437–439 (2003).
42. Humayun, M., Qin, L. P. & Norman, M. D. Geochemical evidence for excess iron in the mantle beneath Hawaii. *Science* **306**, 91–94 (2004).
43. Holzapfel, C., Chakraborty, S., Rubie, D. C. & Frost, D. J. Effect of pressure on Fe–Mg, Ni and Mn diffusion in $(\text{Fe}, \text{Mg}_{1-x})_2\text{SiO}_4$ olivine. *Phys. Earth Planet. Inter.* **162**, 186–198 (2007).
44. Kelemen, P. B., Hart, S. R. & Bernstein, S. Silica enrichment in the continental upper mantle via melt/rock reaction. *Earth Planet. Sci. Lett.* **164**, 387–406 (1998).
45. Mangilik, A. & Christensen, U. R. Effect of lithospheric root on decompression melting in plume–lithosphere interaction models. *Geophys. J. Int.* **164**, 259–270 (2006).
46. Ryabchikov, I. D., Solovova, I. P., Ntafos, T., Buchl, A. & Tikhonov, P. I. Subalkaline picobasalts and plateau basalts from the Putorana plateau (Siberian continental flood basalt province). II. Melt inclusion chemistry, composition of ‘primary’ magmas and P – T regime at the base of the superplume. *Geochem. Int.* **39**, 432–446 (2001).
47. Safonov, O. G., Kamenetsky, V. S. & Perchuk, L. L. Links between carbonatite and kimberlite melts in chloride–carbonate–silicate systems: experiments and application to natural assemblages. *J. Petrol.* **52**, 1307–1331 (2011).
48. Dasgupta, R. & Hirschmann, M. M. The deep carbon cycle and melting in Earth's interior. *Earth Planet. Sci. Lett.* **298**, 1–13 (2010).
49. Sobolev, S. V. & Babeyko, A. Y. What drives orogeny in the Andes? *Geology* **33**, 617–620 (2005).
50. Babeyko, A. Y., Sobolev, S. V., Trumbull, R. B., Oncken, O. & Lavie, L. L. Numerical models of crustal scale convection and partial melting beneath the Altiplano–Puna plateau. *Earth Planet. Sci. Lett.* **199**, 373–388 (2002).
51. Poliakov, A. N., Cundall, P. A., Podladchikov, Y. Y. & Lyakhovsky, V. A. in *Flow and Creep in the Solar System: Observations, Modelling and Theory* (eds Stone, D. B. & Runcorn, S. K.) 175–195 (Kluwer Academic, 1993).
52. Katz, R. F., Spiegelman, M. & Langmuir, C. H. A new parameterization of hydrous mantle melting. *Geochem. Geophys. Geosyst.* **4**, doi:10.1029/2002gc000433 (2003).
53. Spandler, C., Yaxley, G., Green, D. H. & Rosenthal, A. Phase relations and melting of anhydrous K-bearing eclogite from 1200 to 1600 degrees C and 3 to 5 GPa. *J. Petrol.* **49**, 771–795 (2008).
54. Kelemen, P. B., Hirth, G., Shimizu, N., Spiegelman, M. & Dick, H. J. B. A review of melt migration processes in the adiabatically upwelling mantle beneath oceanic spreading ridges. *Phil. Trans. R. Soc. Lond. A* **355**, 283–318 (1997).
55. Dasgupta, R. & Hirschmann, M. M. Effect of variable carbonate concentration on the solidus of mantle peridotite. *Am. Mineral.* **92**, 370–379 (2007).
56. Wignall, P. B. *et al.* Volcanism, mass extinction and carbon isotope fluctuations in the Middle Permian of China. *Science* **324**, 1179–1182 (2009).
57. Retallack, G. J. & Jahren, A. H. Methane release from igneous intrusion of coal during late Permian extinction events. *J. Geol.* **116**, 1–20 (2008).

The evolution of overconfidence

Dominic D. P. Johnson¹ & James H. Fowler²

Confidence is an essential ingredient of success in a wide range of domains ranging from job performance and mental health to sports, business and combat^{1–4}. Some authors have suggested that not just confidence but overconfidence—believing you are better than you are in reality—is advantageous because it serves to increase ambition, morale, resolve, persistence or the credibility of bluffing, generating a self-fulfilling prophecy in which exaggerated confidence actually increases the probability of success^{3–8}. However, overconfidence also leads to faulty assessments, unrealistic expectations and hazardous decisions, so it remains a puzzle how such a false belief could evolve or remain stable in a population of competing strategies that include accurate, unbiased beliefs. Here we present an evolutionary model showing that, counterintuitively, overconfidence maximizes individual fitness and populations tend to become overconfident, as long as benefits from contested resources are sufficiently large compared with the cost of competition. In contrast, unbiased strategies are only stable under limited conditions. The fact that overconfident populations are evolutionarily stable in a wide range of environments may help to explain why overconfidence remains prevalent today, even if it contributes to hubris, market bubbles, financial collapses, policy failures, disasters and costly wars^{9–13}.

Humans show many psychological biases, but one of the most consistent, powerful and widespread is overconfidence. Most people show a bias towards exaggerated personal qualities and capabilities, an illusion of control over events, and invulnerability to risk (three phenomena collectively known as ‘positive illusions’)^{2–4,14}. Overconfidence amounts to an ‘error’ of judgement or decision-making, because it leads to overestimating one’s capabilities and/or underestimating an opponent, the difficulty of a task, or possible risks. It is therefore no surprise that overconfidence has been blamed throughout history for high-profile disasters such as the First World War, the Vietnam war, the war in Iraq, the 2008 financial crisis and the ill-preparedness for environmental phenomena such as Hurricane Katrina and climate change^{9,12,13,15,16}.

If overconfidence is both a widespread feature of human psychology and causes costly mistakes, we are faced with an evolutionary puzzle as to why humans should have evolved or maintained such an apparently damaging bias. One possible solution is that overconfidence can actually be advantageous on average (even if costly at times), because it boosts ambition, morale, resolve, persistence or the credibility of bluffing. If such features increased net payoffs in competition or conflict over the course of human evolutionary history, then overconfidence may have been favoured by natural selection^{5–8}.

However, it is unclear whether such a bias can evolve in realistic competition with alternative strategies. The null hypothesis is that biases would die out, because they lead to faulty assessments and suboptimal behaviour. In fact, a large class of economic models depend on the assumption that biases in beliefs do not exist¹⁷. Underlying this assumption is the idea that there must be some evolutionary or learning process that causes individuals with correct beliefs to be rewarded (and thus to spread at the expense of individuals with incorrect beliefs). However, unbiased decisions are not necessarily the best strategy for

maximizing benefits over costs, especially under conditions of competition, uncertainty and asymmetric costs of different types of error^{8,18–21}. Whereas economists tend to posit the notion of human brains as general-purpose utility maximizing machines that evaluate the costs, benefits and probabilities of different options on a case-by-case basis, natural selection may have favoured the development of simple heuristic biases (such as overconfidence) in a given domain because they were more economical, available or faster.

Here we present a model showing that, under plausible conditions for the value of rewards, the cost of conflict, and uncertainty about the capability of competitors, there can be material rewards for holding incorrect beliefs about one’s own capability. These adaptive advantages of overconfidence may explain its emergence and spread in humans, other animals or indeed any interacting entities, whether by a process of trial and error, imitation, learning or selection. The situation we model—a competition for resources—is simple but general, thereby capturing the essence of a broad range of competitive interactions including animal conflict, strategic decision-making, market competition, litigation, finance and war.

Suppose a resource r is available to an individual that claims it, and there are two individuals, i and j . These individuals each have initial ‘capability’ θ_i and θ_j that determine whether or not they would win a conflict over the resource. Without loss of generality, we assume that θ is distributed in the population according to a symmetric stable probability density²² with cumulative distribution Φ , a mean of 0, and a variance of 0.5. The initial advantage to individual i is $a = \theta_i - \theta_j$, and assumptions about the distribution of θ imply that the probability density of a has a cumulative distribution Φ , a mean of 0, and unit variance (see Supplementary Information for the full model).

If neither individual claims the resource, no fitness is gained. If only one makes a claim, then the claimant acquires the resource and gains fitness r and the other individual gains nothing. If both claim the resource, then both pay a cost c as a result of the conflict between them, but the individual with the higher initial capability will win the conflict, acquiring the resource and obtaining fitness r . This means there are only three outcomes that have an impact on an individual’s fitness: winning a conflict (W), losing a conflict (L), and obtaining an unclaimed resource (O). Given the probability of each of these outcomes (p_W , p_L and p_O), the benefits of obtaining the resource r , and the costs of conflict c , the expected fitness is $E(f) = p_W(r - c) + p_L(-c) + p_O(r)$. Note that r and c can denote expected benefits and costs—if conflict outcomes were made probabilistic instead of deterministic, the results would not change.

Individuals choose whether or not to claim a resource on the basis of their perceived capability relative to the capability of other claimants. If there were no uncertainty in this assessment, there would never be a conflict because the dispute can be settled without cost (the stronger individual takes the resource, and the weaker individual surrenders it, allowing both agents to avoid c)^{23–26}. In the real world, however, uncertainty is common. We therefore model an individual’s uncertainty about his or her opponent’s capability by adding an error term v to the opponent’s capability such that individual i thinks the capability of individual j is $\theta_j + v_i$. To derive analytical results, we assume that this perception error has a magnitude of $\varepsilon > 0$ and is binomially distributed,

¹Politics and International Relations, University of Edinburgh, Edinburgh EH8 9LD, UK. ²Division of Medical Genetics and Department of Political Science, University of California, San Diego, California 92093, USA.

with $\Pr(v = \varepsilon) = \Pr(v = -\varepsilon) = 0.5$ (the 'binomial model'). To evaluate the role of confidence, we allow individuals to perceive their own capability as $\theta + k$, where $k = 0$ indicates unbiased individuals who perceive their capability correctly, $k > 0$ indicates overconfident individuals who think they are stronger than they actually are, and $k < 0$ indicates underconfident individuals who think they are weaker than they actually are.

We explore the emergence and stability of biases in hypothetical populations by using standard assumptions about evolutionary dynamics²⁷ under which the fittest are more likely to survive or reproduce, or the less fit are more likely to copy better strategies. Figure 1a shows regions of the parameter space and five equilibria that occur in the binomial model, all confirmed both analytically and numerically (see Supplementary Information).

When $r/c > 3/2$, the unique equilibrium is a pure (monomorphic) population of overconfident individuals, all of whom evolve a level of overconfidence that is equal to the size of the perception error ($k^* = \varepsilon$). As long as there is at least some perception error, overconfident individuals resist invasion by all other individuals, including underconfident ($k < 0$), unbiased ($k = 0$) and other kinds of overconfident individuals ($k > 0$).

When $1/3 < r/c < 3/2$, there are two equilibria. First, a mixed (polymorphic) population made up of overconfident individuals ($k^* = \varepsilon$) and underconfident individuals ($k^* = -\varepsilon$) is always possible

as long as there is at least some perception error. Second, an unbiased equilibrium ($k^* = 0$) is also possible in this region, but only if the perception error is sufficiently low.

Finally, when $r/c < 1/3$ there are two more equilibria. A pure equilibrium of underconfident individuals ($k^* = -\varepsilon$) is always possible, and a mixed equilibrium of very underconfident ($k^* = -2\varepsilon$) and unbiased ($k^* = 0$) individuals is possible when there is a moderate amount of uncertainty.

The underlying assumptions of the binomial model are deliberately simple to make closed-form characterizations tractable. We also used numerical simulation methods to evaluate the model when we allow the perception error v to vary continuously, using a normal distribution with mean 0 and standard deviation ε (the 'normal model'; Supplementary Information). This assumption may be more realistic than the binomial assumption because it allows perception errors to vary in magnitude.

As with the binomial model, the normal model shows that overconfidence ($k^* > 0$) is the unique pure equilibrium when the benefit/cost ratio is high enough (roughly $r/c > 0.7$; see Fig. 1b), which is notably less stringent than the binomial model reported above. When the benefit/cost ratio falls below this critical value, the unique pure equilibrium is underconfidence ($k^* < 0$). If there is any perception error whatsoever, an absence of bias is only an equilibrium at a single point—the value of resources and the cost of conflict must be in perfect

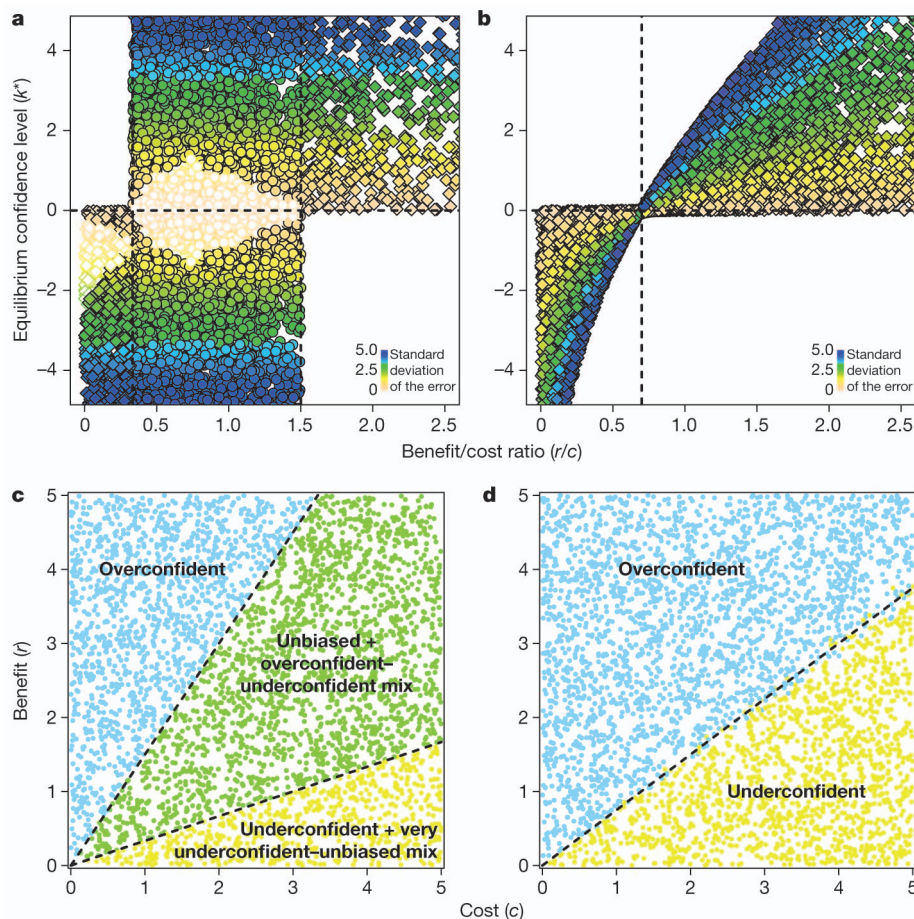


Figure 1 | Best performing levels of confidence across different parameter values. **a, b**, Equilibrium levels of confidence k^* for varying benefit/cost ratios (r/c) and degrees of uncertainty about the capabilities of competitors when assessment errors are modelled with a binomial distribution (**a**) or a normal distribution (**b**). Each point shows the results from a single simulation where the cost, benefit and degree of uncertainty were drawn from a uniform distribution (see Supplementary Information). Each panel shows a total of 10,000 simulations. Shapes indicate types of equilibrium that exist for a given parameter combination:

diamonds, monomorphic; circles, polymorphic (filled shapes indicate that unbiased strategies are not possible). Colours indicate the degree of uncertainty (the standard deviation of the error as defined on the scales). **c, d**, The same results for the binomial (**c**) and normal (**d**) models as a function of costs and benefits (colours indicate what kind of equilibria are possible; these results hold for all levels of perception error). Both models show that overconfident strategies are the unique equilibrium when the benefit/cost ratio is sufficiently high, and unbiased strategies are only possible under limited conditions.

balance to eliminate bias (Fig. 1b). This result suggests that models based on the assumption that individuals perceive their own capabilities without bias¹⁷ are unrealistic: any small change in the benefit/cost ratio will tilt the advantage away from unbiased individuals towards those that assume they are more or less capable than they really are.

The normal model also yields the same positive relationship between perception error and confidence that was derived in the binomial model. As uncertainty about opponent capabilities increases, it becomes more advantageous to express stronger bias (the overconfident become even more confident, and the underconfident become even less confident).

The simulations allowed us to examine some extensions of the model (see Supplementary Information). If we generalize the model to three players, overconfidence is favoured at the same threshold ($r/c > 0.7$). Results are also robust if we allow conflict costs to vary between winners and losers. In fact, the threshold required for overconfidence decreases as losers suffer more. For example, overconfidence evolves when $r/c > 0.6$ if the costs to the winner are $0.8c$, and it evolves when $r/c > 0.45$ if the costs to the winner are $0.2c$. In other words, when conflict for the winner is cheap, overconfidence is even more likely to evolve and persist.

Our model shares interesting parallels with the famous Hawk–Dove game in evolutionary game theory²⁴. ‘Hawks’ escalate until they win (with benefit b) or sustain significant injury (with cost c). ‘Doves’ only display, and retreat if attacked. Where $b > c$, Hawks take over the population and animals always fight. Where $c > b$, a mixed population of Hawks and Doves emerges. The Hawk–Dove game is important because it shows that (where $c > b$) contests can be resolved by ‘conventional’ signals (displays only) with minimal fighting—explaining why many animals have dangerous weapons (such as sharp horns or teeth) but death is rare.

We find that the Hawk–Dove game is a special case of our model, in which the only possible strategies are to be infinitely overconfident ($k = \infty$; that is, Hawk) and therefore always claim the resource, or infinitely underconfident ($k = -\infty$; that is, Dove) and therefore never claim. As we show (see Supplementary Information), the standard equilibria of the Hawk–Dove game emerge under these conditions. Strikingly, however, somewhat overconfident (but not infinitely overconfident) individuals always beat both Hawk and Dove. Our model therefore shows that individuals with a more nuanced strategy—even a biased one—do better than the ‘extreme’ strategies of Hawk and Dove. Moreover, hawkish (overconfident) strategies can dominate even where $c > r$, a finding that contrasts with previous Hawk–Dove models.

Another important result of the model is that environments with more valuable resources will generate more conflict (see Supplementary Information). This parallels the finding in the literature on animal fighting that, where very valuable resources are at stake, hawkish strategies become more common and, in contrast with much animal conflict that is ritualized and restrained, fighting under these conditions can become lethal²⁸.

The analysis here demonstrates that overconfidence often prevails over accurate assessment. Overconfidence is advantageous because it encourages individuals to claim resources they could not otherwise win if it came to a conflict (stronger but cautious rivals will sometimes fail to make a claim), and it keeps them from walking away from conflicts they would surely win. These results conform with previous observations that systematic overestimates of the probability of winning simple gambling games can be adaptive if the benefits of the resource at stake sufficiently exceed the costs of attempting to gain it^{19,20}, that aggressive strategies (such as ‘Hawk’ in Hawk–Dove games) are favoured if the advantages of winning exceed the costs of injury²⁴, and that overconfident states can outperform others in an agent-based model of conflict²⁹.

Note that overconfidence in our model is purely self-deception—there is no other-deception (‘bluffing’) because there is no signalling of k (opponents are not gullible to others’ inflated beliefs). This is important because it demonstrates that there are adaptive advantages

of overconfidence irrespective of any possible (additional) advantages of bluffing. Bluffing is often argued to be unstable in nature because there would be strong selection on discriminating responses. However, this may be partly why self-deception evolved: ‘hiding the truth from yourself to hide it more deeply from others’^{6,7}. Previous work has also shown that bluffing can survive counter-selection if there is ambiguity in one’s own or others’ strengths. If so, bluffs and reality cannot be reliably distinguished, and calling another’s bluff takes on a cost of its own. It has been suggested²⁴ that bluffing is therefore more likely (even if it is detectable in principle) among animals in which serious injury is possible—that is, those with weapons—because the costs of calling a bluff can be high.

Our model applies to any replicating entity or any species, but it has particular implications for humans. First, if contested resources were sufficiently valuable compared with the costs of competing for them during human evolutionary history, we might expect humans to have evolved a bias towards overconfidence^{5,12,19,20}. Such an outcome is exactly what the literature on experimental psychology has long demonstrated but has lacked an explanation for its origin^{2–4,14}. A recent review of whether any ‘false beliefs’ could be biologically adaptive concluded that there is just a single compelling candidate: positive illusions⁸. Today, we may retain evolved proximate mechanisms that give rise to overconfidence even in situations in which the costs of conflict have increased relative to the value of the reward, making overconfidence maladaptive in many modern settings (such as, perhaps, in interpersonal aggression and war).

Second, overconfidence can arise and spread more quickly among humans than other organisms. Rather than relying on genetic mutation and natural selection over many generations, overconfidence in humans can emerge and spread much more rapidly by other means such as trial and error, imitation or learning (which may also generate considerable variation among different ‘ecological’ contexts such as habitats, cultures or organizations). These processes of cultural selection may affect how different levels of confidence emerge, survive and spread today among interacting entities, whether individuals, groups, negotiators, lawyers, traders, banks, sports teams, firms, armies or states. In many of these settings, overconfidence may be beneficial on average even though it only attracts attention when it causes costly disasters, or when the environment (the ratio r/c) changes such that overconfidence begins to generate net costs.

Other recent models have explored the evolution of risk preferences³⁰; however, in the present model, individuals do not prefer or avoid risk—their heuristic is simply to assess capabilities and claim the resource if they perceive a capability gap. As we show (see Supplementary Information), this heuristic causes individuals to behave as though they were calculating the expected outcome of a risky choice under a specific set of assumptions about themselves and their opponents and comparing it with a required risk premium, which is cognitively a much more demanding task. Thus, although it is possible that risk preferences contribute to behaviour in competition and conflict, the simpler mechanism of overconfidence provides a short-cut that yields equivalent outcomes. Such short-cuts may have been favoured in our evolution because they have lower operating costs, were more easily available to natural selection or are capable of reaching decisions faster. In fact, there are many examples of biases in human judgement and decision-making that seem to be adaptive precisely because they offer simple heuristics that deceive us into fitness-maximizing behaviour^{18,20}.

The finding that the optimal level of bias increases with the magnitude of uncertainty is especially intriguing. It suggests that we should expect extreme levels of overconfidence (hubris) or underconfidence (fear) precisely when we are dealing with unfamiliar or poorly understood strategic contexts. We predict that where the value of a prize sufficiently exceeds the costs of competing, overconfidence will be particularly prevalent in some very important domains that have inherently high levels of uncertainty, including international relations (where events are complex and distant and involve foreign cultures

and languages), rare or unpredictable phenomena (such as natural disasters and climate change), novel or complex technologies (such as the Internet bubble and modern financial instruments) and new and untested leaders, allies and enemies. Although overconfidence may have been adaptive in our past, and may still be adaptive in some settings today, it seems that we are likely to become overconfident in precisely the most dangerous of situations.

Received 27 May; accepted 25 July 2011.

- Kanter, R. M. *Confidence: How Winning Streaks and Losing Streaks Begin and End* (Crown Business, 2004).
- Taylor, S. E. & Brown, J. D. Positive illusions and well-being revisited: separating fact from fiction. *Psychol. Bull.* **116**, 21–27 (1994).
- Taylor, S. E. *Positive Illusions: Creative Self-Deception and the Healthy Mind* (Basic Books, 1989).
- Peterson, C. *A Primer in Positive Psychology* (Oxford Univ. Press, 2006).
- Wrangham, R. W. Is military incompetence adaptive? *Evol. Hum. Behav.* **20**, 3–17 (1999).
- Trivers, R. L. The elements of a scientific theory of self-deception. *Ann. NY Acad. Sci.* **907**, 114–131 (2000).
- Trivers, R. *Deceit and Self-Deception: Fooling Yourself the Better to Fool Others* (Allen Lane, 2011).
- McKay, R. T. & Dennett, D. C. The evolution of misbelief. *Behav. Brain Sci.* **32**, 493–510 (2009).
- Tuchman, B. W. *The March of Folly: From Troy to Vietnam* (Alfred A. Knopf, 1984).
- Camerer, C. & Lovallo, D. Overconfidence and excess entry: an experimental approach. *Am. Econ. Rev.* **89**, 306–318 (1999).
- Malmendier, U. & Tate, G. CEO overconfidence and corporate investment. *J. Finance* **60**, 2661–2700 (2005).
- Johnson, D. D. P. *Overconfidence and War: The Havoc and Glory of Positive Illusions* (Harvard University Press, 2004).
- Johnson, D. D. P. & Tierney, D. R. The Rubicon theory of war: how the path to conflict reaches the point of no return. *Int. Secur.* **36**, 7–40 (2011).
- Sharot, T. *The Optimism Bias: A Tour of The Irrationally Positive Brain* (Pantheon, 2011).
- Johnson, D. D. P. & Levin, S. A. The tragedy of cognition: psychological biases and environmental inaction. *Curr. Sci.* **97**, 1593–1603 (2009).
- Akerlof, G. A. & Shiller, R. J. *Animal Spirits: How Human Psychology Drives the Economy, and Why it Matters for Global Capitalism* (Princeton Univ. Press, 2009).
- Fudenberg, D. & Tirole, J. Perfect Bayesian equilibrium and sequential equilibrium. *J. Econ. Theory* **53**, 236–260 (1991).
- Gigerenzer, G. *Adaptive Thinking: Rationality in the Real World* (Oxford Univ. Press, 2002).
- Nettle, D. in *Emotion, Evolution and Rationality* (eds Evans, D. & Cruse, P.) 193–208 (Oxford Univ. Press, 2004).
- Haselton, M. G. & Nettle, D. The paranoid optimist: an integrative evolutionary model of cognitive biases. *Pers. Soc. Psychol. Rev.* **10**, 47–66 (2006).
- Cosmides, L. & Tooby, J. Better than rational: evolutionary psychology and the invisible hand. *Am. Econ. Rev.* **84**, 327–332 (1994).
- Fama, E. F. & Roll, R. Some properties of symmetric stable distributions. *J. Am. Stat. Assoc.* **63**, 817–836 (1968).
- Fearon, J. D. Rationalist explanations for war. *Int. Organ.* **49**, 379–414 (1995).
- Maynard Smith, J. & Parker, G. The logic of asymmetric contests. *Anim. Behav.* **24**, 159–175 (1976).
- Parker, G. A. Assessment strategy and the evolution of fighting behaviour. *J. Theor. Biol.* **47**, 223–243 (1974).
- Enquist, M. & Leimar, O. Evolution of fighting behaviour: decision rules and assessment of relative strength. *J. Theor. Biol.* **102**, 387–410 (1983).
- Nowak, M. A. *Evolutionary Dynamics: Exploring the Equations of Life* (Belknap Press, 2006).
- Enquist, M. & Leimar, O. The evolution of fatal fighting. *Anim. Behav.* **39**, 1–9 (1990).
- Johnson, D. D. P., Weidmann, N. B. & Cederman, L.-E. Fortune favours the bold: an agent-based model reveals adaptive advantages of overconfidence in war. *PLoS ONE* **6**, e20851 (2011).
- McDermott, R., Fowler, J. H. & Smirnov, O. On the evolutionary origin of prospect theory preferences. *J. Polit.* **70**, 335–350 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Barrett, D. Blumstein, L.-E. Cederman, D. Fessler, P. Gočev, M. Haselton, D. Nettle, J. Orbell, K. Panchanathan, M. Price, D. Tierney, R. Trivers, N. Weidmann and R. Wrangham for discussions and help leading to this paper.

Author Contributions D.J. and J.F. conceived the study. J.F. performed the modelling. D.J. and J.F. analysed the results, revised the models and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.J. (dominic.johnson@ed.ac.uk).

Parallel evolution of domesticated *Caenorhabditis* species targets pheromone receptor genes

Patrick T. McGrath¹, Yifan Xu¹, Michael Ailion², Jennifer L. Garrison¹, Rebecca A. Butcher³ & Cornelia I. Bargmann¹

Evolution can follow predictable genetic trajectories¹, indicating that discrete environmental shifts can select for reproducible genetic changes^{2–4}. Conspecific individuals are an important feature of an animal's environment, and a potential source of selective pressures. Here we show that adaptation of two *Caenorhabditis* species to growth at high density, a feature common to domestic environments, occurs by reproducible genetic changes to pheromone receptor genes. Chemical communication through pheromones that accumulate during high-density growth causes young nematode larvae to enter the long-lived but non-reproductive dauer stage. Two strains of *Caenorhabditis elegans* grown at high density have independently acquired multigenic resistance to pheromone-induced dauer formation. In each strain, resistance to the pheromone ascaroside C3 results from a deletion that disrupts the adjacent chemoreceptor genes *serpentine receptor class g* (*srg*)-36 and -37. Through misexpression experiments, we show that these genes encode redundant G-protein-coupled receptors for ascaroside C3. Multigenic resistance to dauer formation has also arisen in high-density cultures of a different nematode species, *Caenorhabditis briggsae*, resulting in part from deletion of an *srg* gene paralogous to *srg*-36 and *srg*-37. These results demonstrate rapid remodelling of the chemoreceptor repertoire as an adaptation to specific environments, and indicate that parallel changes to a common genetic substrate can affect life-history traits across species.

Caenorhabditis elegans and many other nematode species evaluate environmental conditions to choose between two alternative developmental trajectories, one leading to rapid reproduction and one leading to arrest in the long-lived, stress-resistant dauer larva stage. High population density, limiting food and high temperature promote dauer larva formation⁵ (Fig. 1a), a stage that corresponds to the infectious juvenile stage of parasitic nematodes. Dauer larvae do not feed or reproduce, but can survive under conditions that kill other stages, and respond to environmental improvements by exiting the dauer stage and resuming reproductive development. Although the pheromone cues that signal nematode density are normally integrated with food availability, pheromone accumulation in high-density liquid cultures causes animals to form dauer larvae despite the presence of ample food⁶. Non-reproducing dauer animals would seem to be at a disadvantage relative to those that continue to grow in these conditions. To examine adaptation to high-density culture conditions, we measured dauer formation in two laboratory strains of *C. elegans*, LSJ2 and CC1, that were grown in liquid axenic media for approximately 50 years and 4 years, respectively, before permanent cultures were frozen down^{7,8} (Fig. 1b and Methods). Unlike wild-caught strains⁹ and the standard laboratory strain N2 (ref. 10), which readily form dauers in response to partially purified N2 dauer pheromone, CC1 and LSJ2 strains formed almost no dauer larvae (Fig. 1c).

N2, LSJ2 and CC1 arose from a common, inbred *C. elegans* ancestor after isolation from the wild (Fig. 1b), so the pheromone resistance of LSJ2 and CC1 strains must result from new mutations that occurred in the laboratory. The genetic basis of dauer pheromone resistance was

characterized by generating 94 recombinant inbred lines (RILs) between LSJ2 and N2 (Supplementary Fig. 1) that were genotyped at 176 informative single nucleotide polymorphisms (SNPs) (Supplementary Table 1) identified by whole-genome sequencing of LSJ2 and N2 strains (Supplementary Tables 2 and 3). Initial genetic mapping of dauer formation using N2-derived dauer pheromone preparations and the N2–LSJ2 RILs indicated that the trait was multigenic (data not shown). The active components of dauer pheromone are ascarosides, a group of small molecules with a common sugar scaffold and variable side chains^{11–13}. Four individual ascarosides that effectively induced N2 dauer formation (C3, C5, C6 and C9) did not induce dauer formation in LSJ2 or CC1 (Fig. 1c). To simplify trait-mapping, we examined dauer formation in response to individual ascarosides, focusing on the C3 ascaroside, whose receptors and cellular sites of action are unknown. Among 16 RILs exposed to 1 μ M C3, eight formed dauers at a rate comparable to N2 and eight formed dauers at a rate comparable to LSJ2 (Fig. 1d). This bimodal distribution indicates the existence of a single locus that confers C3 resistance.

Quantitative trait locus (QTL) mapping using these sixteen RILs identified a single region on the X chromosome that correlated with the C3 response (Fig. 1e). Mapping of the X-linked C3 resistance locus was verified by creating a near-isogenic line (NIL) with the candidate region from LSJ2 introgressed into an N2 background by ten generations of backcrossing (Fig. 1f). The LSJ2–N2 NIL was resistant to dauer formation induced by C3 ascaroside across a broad range of concentrations (Fig. 1g). Unlike the parental LSJ2 strain, the LSJ2–N2 NIL formed dauer larvae in the presence of three other ascarosides (Fig. 1g). These results identify an X-linked C3-resistance locus as one of several loci that confer pheromone resistance on LSJ2.

To identify the genetic changes in LSJ2 associated with C3 resistance, we sequenced the LSJ2 and N2 strains and identified all fixed polymorphisms between the two strains (Supplementary Tables 2 and 3). The region of the X chromosome associated with C3 resistance included four SNPs in intronic or intergenic regions and a deletion of 4,906 base pairs (bp) in the LSJ2 strain that disrupts two predicted G-protein-coupled receptor genes (*srg*-36 and *srg*-37) (Fig. 2a). A genomic clone from the N2 strain that contains both *srg*-36 and *srg*-37 fully rescued C3 resistance when introduced into the LSJ2–N2 NIL strain, indicating that this deletion causes the C3 resistance associated with the X-linked QTL (Fig. 2b). Notably, *srg*-36 and *srg*-37 were also disrupted by a 6,795-bp deletion in the CC1 strain (Fig. 2a). The deletions in CC1 and LSJ2 have different breakpoints, indicating that they occurred independently. To ask whether deletion of *srg*-36 and *srg*-37 also caused resistance to C3 in CC1, the region surrounding the *srg*-36 and *srg*-37 deletion was introgressed from CC1 into N2 to make a CC1–N2 NIL strain (Fig. 1f). The CC1–N2 NIL was resistant to dauer formation induced by C3 ascaroside (Fig. 2b) and its C3-resistance phenotype was rescued by a transgene covering the *srg*-36 and *srg*-37 genomic regions (Fig. 2b). The CC1–N2 NIL readily formed dauers in response to other ascarosides (Supplementary Fig. 2), indicating that additional genetic mutations contribute to pheromone resistance in

¹Howard Hughes Medical Institute, Laboratory of Neural Circuits and Behavior, The Rockefeller University, New York, New York 10065, USA. ²Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA. ³Department of Chemistry, University of Florida, Gainesville, Florida 32611, USA.

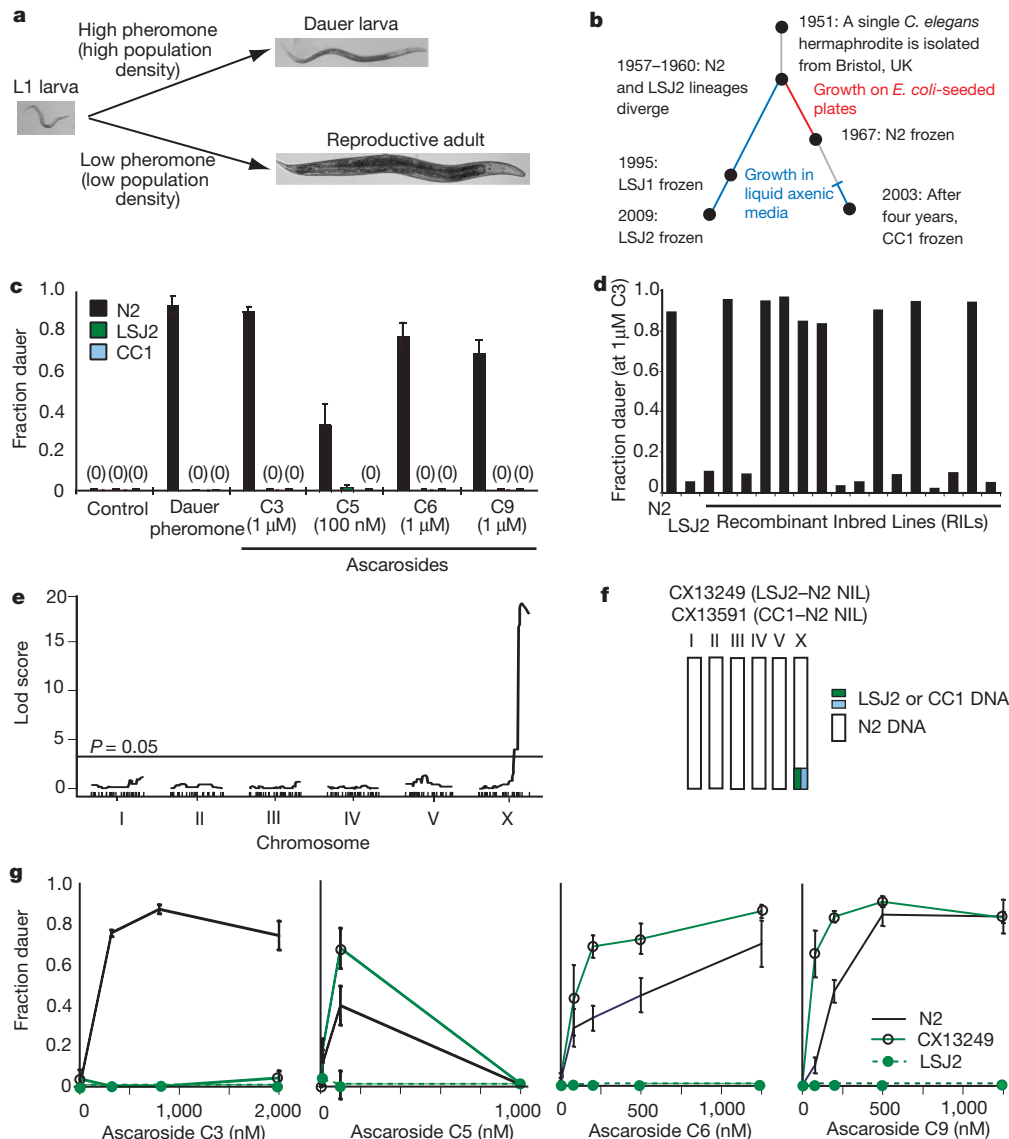


Figure 1 | Strains of *C. elegans* cultivated in liquid are resistant to dauer pheromones. **a**, The developmental decision between reproductive growth and dauer larva formation is regulated by temperature, food and population density. Population density is assessed by the release and sensation of ascarosides including C3, C5, C6 and C9. **b**, History of the *C. elegans* strains N2, LSJ1, LSJ2 and CC1 (see Methods). **c**, Dauer formation of N2, LSJ2 and CC1 in

response to crude dauer pheromone or synthetic ascarosides. **d**, Dauer formation in response to synthetic C3 ascaroside. **e**, QTL mapping of C3 resistance. **f**, Schematic of NILs with a small region from LSJ2 or CC1 introgressed into N2. **g**, Dauer formation in N2, LSJ2 and CX13249 strains. Error bars represent s.e.m.

CC1. The existence of independent C3-resistance mutations affecting *srg-36* and *srg-37* in LSJ2 and CC1 provides strong genetic evidence linking these two chemoreceptors to dauer formation.

To determine which of the two predicted genes is associated with C3 sensitivity, we introduced *srg-36* and *srg-37* complementary DNAs (cDNAs) with their respective upstream regions into the C3-resistant LSJ2–N2 NIL strain (Fig. 2a). Transgenic strains expressing either of the two cDNAs formed dauer larvae in response to C3, although the *srg-37* transgene was less active than the *srg-36* transgene (Fig. 2b). An *srg-37* genomic fragment also rescued dauer formation (Fig. 2b). These results indicate that the *srg-36* and *srg-37* genes are at least partially redundant; either can support dauer formation in response to C3 ascaroside.

The expression patterns of *srg-36* or *srg-37* were inferred from bicistronic transcripts expressing green fluorescent protein (GFP) downstream of the *srg-36* or *srg-37* promoter and cDNA. These *srg-36* and *srg-37* reporter transgenes rescued C3-induced dauer formation (Fig. 2b), and were most strongly and consistently expressed in the

ASI chemosensory neurons, with weak or inconsistent expression in a few other neurons (Fig. 2c). Reporters for *srg-36* and *srg-37* were robustly expressed during the L1 stage when the dauer decision is made (Fig. 2c). The ASI neurons are primary regulators of dauer formation¹⁴, and are therefore plausible sites of *srg-36* and *srg-37* action. An *srg-36* cDNA driven by the ASI-selective *srg-47* promoter rescued C3-induced dauer formation in the LSJ2–N2 NIL, but expression of *srg-36* in AFD or ASE sensory neurons did not (Fig. 2b and Supplementary Fig. 3). These results are consistent with the hypothesis that *srg-36* acts in ASI to sense ascaroside C3 (Supplementary Fig. 4).

The subcellular localization of SRG-36 was examined by fusing GFP to the *srg-36* cDNA and expressing the hybrid gene from an ASI-specific promoter. This fusion protein was primarily localized in the sensory cilia of ASI (Fig. 3a), indicating a sensory function for SRG-36. The selective association of *srg-36* and *srg-37* with C3 responsiveness, and not with responsiveness to other ascarosides, suggested that they might encode C3 receptors. We tested this hypothesis by a gain-of-function experiment in which the *srg-36* cDNA or the *srg-37* cDNA

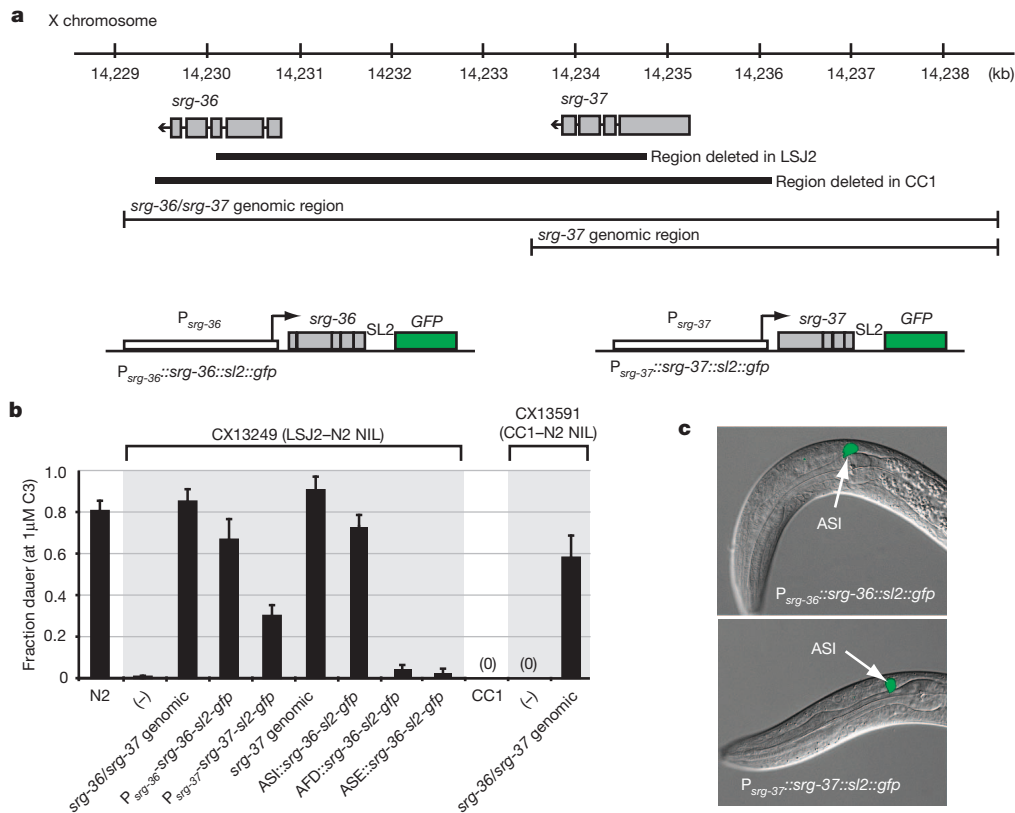


Figure 2 | Resistance to C3 ascaroside is caused by deletion of two *srg* genes. **a**, Genomic region surrounding *srg-36* and *srg-37* on the X chromosome, deletion breakpoints in LSJ2 and CC1 strains, fragments used for transgenic rescue, and design of bicistronic fusion genes. **b**, Transgenic rescue of dauer formation in response to C3 ascaroside. NIL strains used as recipients for rescue

are shown in Fig. 1f. The ASI promoter was *srg-47* (Supplementary Fig. 3), the AFD promoter was *gcy-8* and the ASE promoter was *flp-6*. Error bars represent s.e.m. **c**, Expression of GFP from bicistronic fusion genes for *srg-36* and *srg-37* in L1 larvae, showing predominant expression in ASI sensory neurons.

was expressed in ASH neurons, a pair of polymodal nociceptive neurons that direct rapid avoidance behaviour¹⁵. Unlike control animals, animals expressing the *ASH::srg-36* or *ASH::srg-37* transgene reversed

rapidly in response to 1 μ M C3 in an acute-avoidance assay (Fig. 3b). Neither *ASH::srg-36*, *ASH::srg-37*, nor control animals responded strongly to 1 μ M C6 (Fig. 3b). These results demonstrate that expression of SRG-36 or SRG-37 in ASH is sufficient for C3-specific behavioural responses.

The ASH neurons respond to repulsive stimuli with increases in intracellular calcium that can be monitored using genetically-encoded calcium indicators¹⁶. Animals expressing *srg-36* or *srg-37* in ASH showed rapid, reliable Ca^{2+} increases in response to 1 μ M C3 ascaroside, but not to 1 μ M C6 ascaroside (Fig. 3c); control animals did not respond to either C3 or C6. These results indicate that SRG-36 and SRG-37 are chemoreceptors (or subunits of chemoreceptors) that sense the C3 ascaroside. Although *srg-36* and *srg-37* are normally expressed in ASI, ASI neurons did not respond to C3 with calcium transients (data not shown). Little is known about pheromone signalling pathways in ASI, so the reason for this negative result is unclear.

LSJ2 was originally propagated in the Dougherty laboratory in the 1950s and 1960s to study nutrient requirements for nematode growth. A strain of *C. briggsae*, DR1690, that was grown in the Dougherty laboratory under the same conditions as LSJ2 also acquired resistance to dauer pheromone¹⁷ (Fig. 4a). *C. briggsae* and *C. elegans* are estimated to have diverged 20–30 million years ago¹⁸. The *C. briggsae* genome encodes several genes closely related to *srg-36* and *srg-37*, but does not have one-to-one orthologues of these genes (Fig. 4b). Comparing genomic DNA sequences from DR1690 with the reference *C. briggsae* strain AF16, we discovered a 33-kilobase deletion in DR1690 that disrupts one of the *srg* paralogs, *CBG24690*, and six other genes (Fig. 4c). To determine whether this deletion affects pheromone responses, we created a NIL with the *CBG24690* deletion introgressed into the AF16 reference background (DR1690–AF16 NIL). As previously

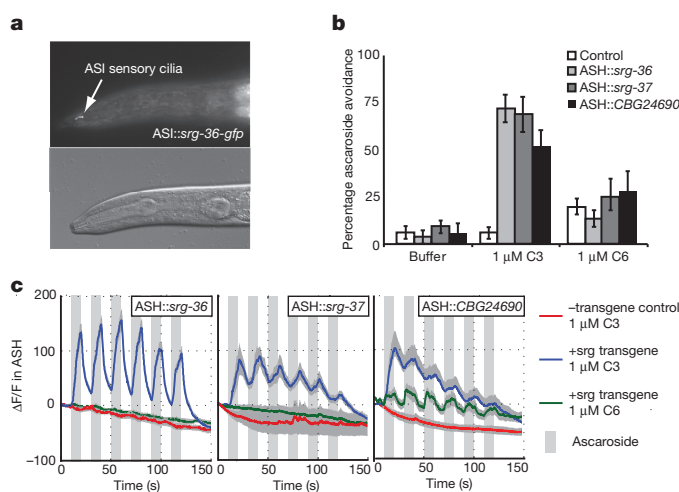


Figure 3 | The *srg* genes encode ascaroside receptors. **a**, Localization of SRG-36::GFP to ASI cilia (L4 animal). **b**, Ascaroside avoidance behaviours of animals with ectopic expression of *srg-36*, *srg-37* or *CBG24690* (shown in Fig. 4) in the ASH nociceptive neurons. Error bars represent s.e.m. **c**, Ascaroside-induced Ca^{2+} transients in ASH neurons that ectopically express *C. elegans* *srg-36* or *srg-37*, or *C. briggsae* *CBG24690*, in ASH. Grey bars indicate the presence of C3 or C6 ascaroside, shading indicates s.e.m., $n \geq 10$ animals per condition. Ca^{2+} was monitored using the genetically-encoded calcium sensor GCaMP3.0 (ref. 30). $\Delta F/F$, percentage fluorescence change (baseline fluorescence = 100%).

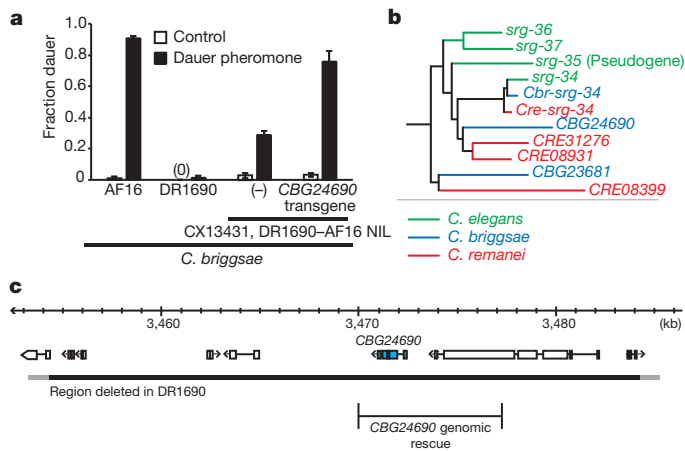


Figure 4 | Evolutionary conservation of *srg* function. **a**, Rescue of *C. briggsae* dauer formation in response to partially purified dauer pheromone, mediated by genomic fragments containing the *CBG24690* gene. CX13431 is a near-isogenic line containing the *CBG24690* deletion from DR1690, introgressed into the AF16 background. Error bars represent s.e.m. **b**, Schematic of genes closely related to *srg-36* and *srg-37* from *C. elegans*, *C. briggsae* and *C. remanei* (adapted from ref. 19). **c**, *CBG24690* genomic region from the AF16 *C. briggsae* reference strain, and location of a large deletion in the DR1690 *C. briggsae* strain that was cultivated for an extended period in liquid axenic media.

reported¹⁷, AF16 readily formed dauers in response to *C. elegans* dauer pheromone, whereas DR1690 animals were resistant to dauer pheromone (Fig. 4a). The *C. briggsae* DR1690–AF16 NIL formed dauers at an intermediate level compared to these two strains (Fig. 4a), indicating that this region contains one of several mutations that contribute to pheromone resistance in DR1690. The DR1690–AF16 NIL was also resistant to purified ascaroside C3 compared to the parental AF16 strain (data not shown). A transgene covering the *CBG24690* genomic region rescued dauer formation in the pheromone-resistant DR1690–AF16 NIL strain (Fig. 4a). These results demonstrate that the *CBG24690* *srg* gene contributes to pheromone-induced dauer formation in *C. briggsae*.

To investigate whether *CBG24690* also encodes an ascaroside receptor, we expressed a *CBG24690* cDNA in the *C. elegans* ASH neurons. Animals expressing the ASH::*CBG24690* transgene reversed rapidly when presented with 1 μ M C3, but not when presented with 1 μ M C6 (Fig. 3b). Animals expressing the *CBG24690* *srg* gene in ASH also showed rapid Ca^{2+} increases in response to 1 μ M C3 (Fig. 3c). Unlike animals expressing *srg-36* or *srg-37* in ASH, animals expressing *CBG24690* in ASH showed weaker but reliable Ca^{2+} increases in response to the related ascaroside C6 (Fig. 3c).

Our results indicate that *srg-36* and *srg-37*, two members of a large nematode-specific family of G-protein-coupled receptors¹⁹, encode redundant receptors for the ascaroside C3. The *srg* gene family is distinct from the *srbc* gene family that was previously implicated in sensing ascarosides C6 and C9 (ref. 20), indicating that at least two of the seven chemoreceptor superfamilies of *C. elegans* can detect ascaroside pheromones. Chemoreceptors are among the fastest-evolving genes in metazoan genomes. They come from entirely different protein families in vertebrates, insects and nematodes, and change rapidly between species²¹: only half of the chemoreceptors in *C. elegans* and *C. briggsae* are one-to-one orthologue pairs¹⁹. Despite the rapid evolution of these genes, the function of *srg*-like genes in pheromone detection has been conserved since *C. briggsae* and *C. elegans* diverged: *srg-36* and *srg-37* in *C. elegans*, and *CBG24690* in *C. briggsae*, each sense C3 ascaroside to induce dauer formation. Some differences between species exist, however, because *CBG24690* also senses C6 ascaroside at concentrations that are not sensed by *srg-36* and *srg-37*. Differences in pheromone production and sensation by different *Caenorhabditis* species may allow both species-specific discrimination and general

detection of *Caenorhabditis* species in the vicinity, as is observed with quorum-sensing systems in bacteria²².

These results demonstrate a reproducible change in the chemoreceptor repertoire in response to a discrete environmental shift. During high-density growth in the laboratory, resistance to ascaroside pheromones arose independently in *C. elegans* LSJ2 and CC1, and in *C. briggsae* DR1690, through changes in related *srg* genes. Although numerous single-gene mutations can convey resistance to dauer formation in *C. elegans*²³, deletion of *srg* genes seems to be a favoured route to pheromone resistance in mixed populations. It is possible that specific features of the chromosomal region surrounding *srg-36* and *srg-37* predispose this region to deletion mutations, but these features would also need to be found near the *CBG24690* gene in *C. briggsae*. Alternatively, the spectrum of potential dauer-defective mutants may be constrained because the known single-gene mutants that are resistant to dauer formation have pleiotropic effects on sensory biology, stress resistance and starvation responses that would reduce their fitness in mixed cultures²³. Global analysis of functional genetic variants indicates that evolutionarily relevant mutations are not randomly distributed, but rather cluster in specific genetic loci, or hotspot genes¹. These observations indicate that genetic trajectories during evolution are constrained and that adaptation can, at least to some extent, be predictable. One class of known adaptive genes are input–output genes, developmental regulators with complex *cis*-regulatory motifs that provide a molecular substrate that allows sculpting of developmental patterns²⁴. The genes *srg-36* and *srg-37* seem to fall into a second class of adaptive genes, including opsin genes and taste receptors^{25,26}: sensory receptors whose diversity allows circumscribed adaptation to environmental changes without pleiotropic effects.

METHODS SUMMARY

The LSJ2 strain and the N2 laboratory strain are descended from one ancestral hermaphrodite isolated by W. Nicholas. LSJ2 was grown continuously in liquid axenic media starting in about 1957 at the Kaiser Foundation Research Institute, The University of California, Berkeley, and San Jose State University, until a sample was frozen in 2009.

Dauer formation assays were performed with crude or synthesized ascarosides as described¹¹. Values report the average fraction of dauer animals 72 h after eggs were laid on assay plates. With the exception of the mapping experiments, each strain and condition was tested in a minimum of five independent assays.

RILs were generated from reciprocal crosses between LSJ2 and an N2-derived strain, and were inbred for ten generations. Two laboratory-derived polymorphisms that modify the *npr-1* and *glb-5* genes in N2 affect many *C. elegans* behaviours^{7,27}; to eliminate their effects, the cross was initiated with a strain containing 99% N2 DNA but the ancestral alleles of *npr-1* and *glb-5* from the CB4856 strain (Supplementary Fig. 1). These RILs were genotyped at 192 SNPs between LSJ2 and N2. The fraction of animals forming dauers in response to 1 μ M C3 ascaroside was used as a phenotype for nonparametric QTL mapping.

The genes *srg-36*, *srg-37* and *CBG24690* were ectopically expressed in ASH using the *sra-6* promoter. Vehicle control, 1 μ M of C3 or 1 μ M of C6 were dissolved in M13 buffer and presented to animals using the drop test²⁸. Each animal was scored three times for the ability to reverse in response to the stimulus. At least 50 animals were scored blindly for each strain and condition.

ASH imaging was performed in a custom-designed microfluidic device²⁹. The genetically-encoded calcium indicator GCaMP3.0 (ref. 30) was expressed in ASH using the *sra-6* promoter.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 February; accepted 20 July 2011.

Published online 17 August 2011.

1. Stern, D. L. & Orgogozo, V. Is genetic evolution predictable? *Science* **323**, 746–751 (2009).
2. Chan, Y. F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**, 302–305 (2010).
3. Protas, M. E. *et al.* Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genet.* **38**, 107–111 (2006).
4. Woods, R., Schneider, D., Winkworth, C. L., Riley, M. A. & Lenski, R. E. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **103**, 9107–9112 (2006).

5. Golden, J. W. & Riddle, D. L. The *Caenorhabditis elegans* dauer larva: developmental effects of pheromone, food, and temperature. *Dev. Biol.* **102**, 368–378 (1984).
6. Stiernagle, T. Maintenance of *C. elegans*. *WormBook* (ed. The *C. elegans* Research Community) doi:10.1895/wormbook.1.101.1 (11 February 2006); available at <http://www.wormbook.org>.
7. McGrath, P. T. *et al.* Quantitative mapping of a digenic behavioral trait implicates globin variation in *C. elegans* sensory behaviors. *Neuron* **61**, 692–699 (2009).
8. Szwedczyk, N. J., Kozak, E. & Conley, C. A. Chemically defined medium and *Caenorhabditis elegans*. *BMC Biotechnol.* **3**, 19 (2003).
9. Viney, M. E., Gardner, M. P. & Jackson, J. A. Variation in *Caenorhabditis elegans* dauer larva formation. *Dev. Growth Differ.* **45**, 389–396 (2003).
10. Golden, J. W. & Riddle, D. L. A pheromone influences larval development in the nematode *Caenorhabditis elegans*. *Science* **218**, 578–580 (1982).
11. Butcher, R. A., Ragains, J. R., Kim, E. & Clardy, J. A potent dauer pheromone component in *Caenorhabditis elegans* that acts synergistically with other components. *Proc. Natl Acad. Sci. USA* **105**, 14288–14292 (2008).
12. Jeong, P. Y. *et al.* Chemical structure and biological activity of the *Caenorhabditis elegans* dauer-inducing pheromone. *Nature* **433**, 541–545 (2005).
13. Srinivasan, J. *et al.* A blend of small molecules regulates both mating and development in *Caenorhabditis elegans*. *Nature* **454**, 1115–1118 (2008).
14. Bargmann, C. I. & Horvitz, H. R. Control of larval development by chemosensory neurons in *Caenorhabditis elegans*. *Science* **251**, 1243–1246 (1991).
15. Kaplan, J. M. & Horvitz, H. R. A dual mechanosensory and chemosensory neuron in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **90**, 2227–2231 (1993).
16. Hilliard, M. A. *et al.* *In vivo* imaging of *C. elegans* ASH neurons: cellular response and adaptation to chemical repellents. *EMBO J.* **24**, 63–72 (2005).
17. Fodor, A., Riddle, D. L., Nelson, F. K. & Golden, J. W. Comparison of a new wild-type *Caenorhabditis briggsae* with laboratory strains of *C. briggsae* and *C. elegans*. *Nematologica* **29**, 203–216 (1983).
18. Cutter, A. D. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Mol. Biol. Evol.* **25**, 778–786 (2008).
19. Thomas, J. H. & Robertson, H. M. The *Caenorhabditis* chemoreceptor gene families. *BMC Biol.* **6**, 42 (2008).
20. Kim, K. *et al.* Two chemoreceptors mediate developmental effects of dauer pheromone in *C. elegans*. *Science* **326**, 994–998 (2009).
21. Nei, M., Niimura, Y. & Nozawa, M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nature Rev. Genet.* **9**, 951–963 (2008).
22. Waters, C. M. & Bassler, B. L. Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* **21**, 319–346 (2005).
23. Hu, P. J. Dauer. *WormBook* (ed. The *C. elegans* Research Community) doi:10.1895/wormbook.1.144.1 (8 August 2007); available at <http://www.wormbook.org>.
24. Sucena, E., Delon, I., Jones, I., Payre, F. & Stern, D. L. Regulatory evolution of *shavenbaby/ovo* underlies multiple cases of morphological parallelism. *Nature* **424**, 935–938 (2003).
25. Kim, U. K. *et al.* Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science* **299**, 1221–1225 (2003).
26. Yokoyama, S. Molecular evolution of vertebrate visual pigments. *Prog. Retin. Eye Res.* **19**, 385–419 (2000).
27. de Bono, M. & Bargmann, C. I. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* **94**, 679–689 (1998).
28. Hilliard, M. A., Bargmann, C. I. & Bazzicalupo, P. C. *C. elegans* responds to chemical repellents by integrating sensory inputs from the head and the tail. *Curr. Biol.* **12**, 730–734 (2002).
29. Chronis, N., Zimmer, M. & Bargmann, C. I. Microfluidics for *in vivo* imaging of neuronal and behavioral activity in *Caenorhabditis elegans*. *Nature Methods* **4**, 727–731 (2007).
30. Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* **6**, 875–881 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Lu for the LSJ2 strain, M. Rockman for a *qgIR1(X,CB4856>N2)* introgression strain, J. Ragains for ascaroside synthesis, H. Hang for assistance in purifying pheromones and S. Dewell, K. Foster, N. Ringstad, A. Bendesky, Y. Saheki, M. Zimmer, S. Crosson, E. Feinberg, E. Toro, M. Rockman and L. Kruglyak for comments and advice. P.T.M. was funded by a Damon Runyon Fellowship. Y.X. was supported by Medical Scientist Training Program (MSTP) grant GM07739 and a Paul and Daisy Soros Fellowship. J.L.G. was an HHMI fellow of the Helen Hay Whitney Foundation and is funded by National Institutes of Health (NIH) K99 GM092859. R.A.B. is supported by R00GM87533. C.I.B. is an investigator of the Howard Hughes Medical Institute. This work was supported by the HHMI.

Author Contributions P.T.M. and C.I.B. designed and interpreted experiments and wrote the paper. P.T.M. performed all genetic, molecular and behavioural experiments, Y.X. conducted calcium imaging experiments, M.A. identified the dauer-formation defect in the LSJ2 lineage, R.A.B. characterized and synthesized ascarosides and J.L.G. contributed reagents.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.I.B. (cori@rockefeller.edu).

METHODS

Strains were cultivated at 22 °C on agar plates seeded with *E. coli* strain OP50.

LSJ2 is a sister strain to the standard N2 laboratory strain. Both LSJ2 and N2 are descended from a single animal isolated by W. Nicholas from a mushroom compost culture provided by L. Staniland in Bristol, England. The strain was transferred to E. Dougherty's laboratory at the Kaiser Foundation Research Institute in the 1950s, and in the late 1950s it separated into two substrains. One of these substrains was mistakenly believed to be *C. briggsae* and represents the LSJ2 lineage. S. Brenner received a cultivar of the second substrain from E. Dougherty in 1964; this cultivar became N2. The LSJ2 lineage was continuously cultivated in liquid axenic media at the University of California, Berkeley and at San Jose State University thereafter. In 1995, a cultivar of the strain was sent to the *Caenorhabditis* genetics centre and frozen to become LSJ1. In 2009, N. Lu from San Jose State University provided a second cultivar of the strain that had been grown for an additional 14 years in axenic media; this strain is designated LSJ2.

Other strains used in this study are: N2, CCI, LSJ1, MY14, AF16, DR1690, CX12311 *kyIR1(V, CB4856>N2)*; *qgIR1(X, CB4856>N2)*, CX13249 *kyIR88(X, LSJ2>N2)*, CX13330 *kyIR88(X, LSJ2>N2)*; *kyEx3927 (srg-36/srg-37 genomic region + P_{elt-2::gfp})*, CX13331 *kyIR88(X, LSJ2>N2)*; *kyEx3928 (srg-36/srg-37 genomic region + P_{elt-2::gfp})*, CX13332 *kyIR88(X, LSJ2>N2)*; *kyEx3929 (srg-37 genomic region + P_{elt-2::gfp})*, CX13333 *kyIR88(X, LSJ2>N2)*; *kyEx3930 (srg-37 genomic region + P_{elt-2::gfp})*, CX13334 *kyIR88(X, LSJ2>N2)*; *kyEx3931 (srg-37 genomic region + P_{elt-2::gfp})*, CX13335 *kyIR88(X, LSJ2>N2)*; *kyEx3932 (P_{srg-36::srg-36:sl2:gfp + P_{elt-2::gfp}})*, CX13336 *kyIR88(X, LSJ2>N2)*; *kyEx3933 (P_{srg-36::srg-36:sl2:gfp + P_{elt-2::gfp}})*, CX13337 *kyIR88(X, LSJ2>N2)*; *kyEx3934 (P_{srg-36::srg-36:sl2:gfp + P_{elt-2::gfp}})*, CX13338 *kyIR88(X, LSJ2>N2)*; *kyEx3935 (P_{srg-37::srg-37:sl2:gfp + P_{elt-2::gfp}})*, CX13339 *kyIR88(X, LSJ2>N2)*; *kyEx3936 (P_{srg-37::srg-37:sl2:gfp + P_{elt-2::gfp}})*, CX13340 *kyIR88(X, LSJ2>N2)*; *kyEx3937 (P_{srg-37::srg-37:sl2:gfp + P_{elt-2::gfp}})*, CX13431 *kyIR94(X, DR1690> AF16)*, CX13591 *kyIR95(X, CCI>N2)*, CX13592 *kyIR95(X, CCI>N2)*; *kyEx4118 (srg-36/srg-37 genomic region + P_{elt-2::gfp})*, CX13593 *kyIR95(X, CCI>N2)*; *kyEx4119 (srg-36/srg-37 genomic region + P_{elt-2::gfp})*, CX13594 *kyIR95(X, CCI>N2)*; *kyEx4120 (srg-36/srg-37 genomic region + P_{elt-2::gfp})*, CX13685 *kyEx2865 (P_{sra-6::GCAMP3.0 + P_{ofm-1::rfp}})*; *kyEx4171 (P_{sra-6::srg-36 + P_{ofm-1::rfp}})*, CX13686 *kyEx2865 (P_{sra-6::GCAMP3.0 + P_{ofm-1::gfp}})*; *kyEx4172 (P_{sra-6::srg-36 + P_{ofm-1::rfp}})*, CX13687 *kyEx2865 (P_{sra-6::GCAMP3.0 + P_{ofm-1::gfp}})*; *kyEx4173 (P_{sra-6::srg-36 + P_{ofm-1::rfp}})*, CX13603 *kyIR88(X, LSJ2>N2)*; *kyEx4125 (P_{sra-6::srg-36:gfp + P_{elt-2::gfp}})*, CX13739 *kyIR94(X, DR1690> AF16)*; *kyEx4201 (CBG24690 genomic region + P_{myo-2::mcherry})*, CX13740 *kyIR94(X, DR1690> AF16)*; *kyEx4202 (CBG24690 genomic region + P_{myo-2::mcherry})*, CX13741 *kyIR94(X, DR1690> AF16)*; *kyEx4203 (CBG24690 genomic region + P_{myo-2::mcherry})*, CX13977 *kyIR88(X, LSJ2>N2)*; *kyEx4316 (P_{srg-47::srg-36:sl2:gfp + P_{ofm-1::rfp}})*, CX13978 *kyIR88(X, LSJ2>N2)*; *kyEx4317 (P_{srg-47::srg-36:sl2:gfp + P_{ofm-1::rfp}})*, CX13979 *kyIR88(X, LSJ2>N2)*; *kyEx4318 (P_{srg-47::srg-36:sl2:gfp + P_{ofm-1::rfp}})*, CX13980 *kyIR88(X, LSJ2>N2)*; *kyEx4319 (P_{srg-47::srg-36:sl2:gfp + P_{ofm-1::rfp}})*, CX13981 *kyIR88(X, LSJ2>N2)*; *kyEx4320 (P_{srg-47::srg-36:sl2:gfp + P_{ofm-1::rfp}})*, CX13982 *kyIR88(X, LSJ2>N2)*; *kyEx4321 (P_{srg-47::srg-36:sl2:gfp + P_{ofm-1::rfp}})*, CX13983 *kyEx2865 (P_{sra-6::GCAMP3.0 + P_{ofm-1::gfp}})*; *kyEx4322 (P_{sra-6::srg-37:c + P_{ofm-1::rfp}})*, CX13984 *kyEx2865 (P_{sra-6::GCAMP3.0 + P_{ofm-1::gfp}})*; *kyEx4323 (P_{sra-6::srg-37:c + P_{ofm-1::rfp}})*, CX13985 *kyEx2865 (P_{sra-6::GCAMP3.0 + P_{ofm-1::gfp}})*; *kyEx4324 (P_{sra-6::srg-37:c + P_{ofm-1::rfp}})*, CX13986 *kyEx2865 (P_{sra-6::GCAMP3.0 + P_{ofm-1::gfp}})*; *kyEx4325 (P_{sra-6::CBG24690 + P_{ofm-1::rfp}})*, CX13987 *kyEx2865 (P_{sra-6::GCAMP3.0 + P_{ofm-1::gfp}})*; *kyEx4326 (P_{sra-6::CBG24690 + P_{ofm-1::rfp}})*, CX13988 *kyEx2865 (P_{sra-6::GCAMP3.0 + P_{ofm-1::gfp}})*; *kyEx4327 (P_{sra-6::CBG24690 + P_{ofm-1::rfp}})*, CX14023 *kyIR88(X, LSJ2>N2)*; *kyEx4342 (P_{flp-6::srg-36:sl2:gfp + P_{ofm-1::rfp}})*, CX14024 *kyIR88(X, LSJ2>N2)*; *kyEx4343 (P_{flp-6::srg-36:sl2:gfp + P_{ofm-1::rfp}})*.

Dauer formation assays. Dauer plates contained 1 µl (for *C. elegans*) or 25 µl (for *C. briggsae*) of crude *C. elegans* dauer pheromone, or 80 nM–2 µM ascarosides (synthesized as previously described^{11,31,32}), in NGM agar without peptone (2.2% Noble Agar, 5 µg ml⁻¹ cholesterol, 15 mM NaCl, 1 mM CaCl₂, 1 mM MgSO₄ and 25 mM KPO₄). For *C. elegans*, 20 µl of heat-killed *E. coli* OP50 bacteria (10 µg ml⁻¹) were added to each plate, five adult animals were picked onto the plate, allowed to lay eggs for 4 h and then removed. Plates were incubated at 25 °C for 72 h before being scored for dauers, identified by a thin body morphology and non-pumping pharynx. At least five plates were assayed for each strain and condition. For *C. briggsae*, OP50 lawns killed with 50 mg ml⁻¹ streptomycin were used, because otherwise animals crawled off the heat-killed bacterial lawn and died. Higher levels of pheromone were required to induce dauer formation on the streptomycin-killed bacteria.

Crude dauer pheromone was purified from 2 l of N2 cultured in S basal medium with HB101 bacteria for 11 days. Supernatants were clarified by centrifugation,

further filtered through a Buchner filter funnel (medium frit, Chemglass) under vacuum, then filtered through 0.2 µm PES membranes (Nalgene), concentrated using a rotary evaporator and lyophilized. Solids were extracted three times with 100% ethanol (100 ml each), and the eluents were combined and concentrated using a rotary evaporator to yield 5 ml of crude dauer pheromone (stored at –20 °C).

LSJ2 and N2 sequencing and analysis. Genomic DNA was isolated from seven strains: LSJ2, LSJ1 (a sample from the LSJ2 lineage frozen in 1995), MY14 (a wild strain used as an outgroup) and four EMS-mutagenized N2-derived strains. Genomic DNA (10 µg) was provided to the Rockefeller Genomics Resource Center for sequencing. DNA samples were processed using the gDNA paired-end sample preparation kit from Illumina, and sequencing was performed using a GAI instrument.

SNP analysis. Sequencing reads with an average quality score above 27 (Sanger format) were aligned to the WS195 *C. elegans* reference sequence and used to identify SNPs using the MAQ software suite (version 0.7.1 easyrun command, using default settings)³³. The final filtered SNPs (the *cns.final.snp* file) for each strain were further analysed using custom software that analysed the number of reference and mutant reads that were present for the polymorphisms in all the sequenced strains. Many of the predicted SNPs, both in LSJ2 and in N2, were supported by reads that matched both the reference N2 nucleotide and a mutant nucleotide. These 'heterozygous' SNPs could represent heterozygous alleles maintained by balancing selection, but different levels of coverage of the two reads indicates that these apparent SNPs are actually alignment errors.

To be considered a true polymorphism between the LSJ2 and the N2 strains, we required at least 90% of the reads from the LSJ2 sequencing to be mutant, and fewer than 10% of the reads from the N2-derived strains to be mutant. A total of 223 SNPs passed these criteria. Using MY14 as an outgroup, the SNPs were then classified into the LSJ2 branch if fewer than 10% of the reads from the MY14 sequencing were mutant, and into the N2 branch if more than 75% of the reads from the MY14 sequencing were mutant. Eight SNPs could not be classified because there were no reads from the MY14 sequencing. We broke down the LSJ2 lineage further into mutations occurring before and after 1995, using sequence from the LSJ1 strain. If more than 90% of the reads from LSJ1 supported the mutant read, then the SNP was classified as occurring before 1995. If fewer than 25% of the reads from LSJ1 supported the mutant read, then the SNP was classified as occurring after 1995. One SNP could not be classified.

A recent whole-genome sequencing report indicated a substantially higher level of mutation between N2 and LSJ1 than we detected here, with 877 SNPs instead of 171 (ref. 34). Fourteen SNPs predicted by that analysis, but not by this one, were examined by PCR and Sanger sequencing of N2 and LSJ1; 13 of the 14 were not confirmed and one was ambiguous. If these SNPs are representative, ~80% of the SNPs in the previous report are either misclassified bases or SNPs specific to that laboratory's strains.

Indel analysis. We created a custom algorithm to identify insertions and deletions (indels) in LSJ2 with respect to the N2 reference. Because MAQ does not use gapped alignment for aligning single-end reads to the reference sequence, we reasoned that most reads covering an insertion or deletion would be unaligned by the MAQ software. We identified regions of low coverage (defined as <12 reads) using custom software and identified any reads unaligned by MAQ with partial matches (defined as reads with 18 contiguous matches) in these regions. We then realigned the partial reads, considering all possible 1-bp insertions and deletions in the low-coverage region. If the 1-bp indel region with the best alignments to the partial matches resulted in an average match of 35 out of 36 bp in all the sequence reads, then we considered this evidence of a real difference from the reference N2 sequence. For each of these 1-bp indels, we searched the unaligned reads from N2 for evidence of an identical polymorphism (again using an average match of 35 out of 36 bp as evidence for the polymorphism), because these 1-bp indels found in both LSJ2 and N2 sequencing are probable reference errors. We considered the remaining 41 1-bp indels as genuine differences between LSJ2 and N2 and classified them into the LSJ2 or N2 lineage using the unaligned MY14 sequencing reads as an outgroup.

The remaining low-coverage regions with partial matches were then visually inspected for the presence of larger deletions or insertions. The exact breakpoints for each deletion or insertion were defined using unambiguous regions, with MY14 as an outgroup, to classify the insertion or deletion into the N2 or LSJ2 lineage. A total of 26 indels larger than 1 bp were identified by this analysis.

A total of 331 indels were identified between LSJ1 and N2 in the previous whole-genome sequencing report³⁴: a significantly higher number than the 67 indels identified here. Unlike the SNPs, we have not assessed the differences in indel predictions by PCR and Sanger sequencing.

Deletion information. Large deletions were verified using Sanger sequencing. The *srg-36/srg-37* region in LSJ2 contained a deletion of 4,906 bp replaced with

relative to F0 was plotted individually for each trial. A second Matlab script was used to plot the average of all trials with standard errors for each time point.

31. Butcher, R. A., Fujita, M., Schroeder, F. C. & Clardy, J. Small-molecule pheromones that control dauer development in *Caenorhabditis elegans*. *Nature Chem. Biol.* **3**, 420–422 (2007).
32. Butcher, R. A., Ragains, J. R. & Clardy, J. An indole-containing dauer pheromone component with unusual dauer inhibitory activity at higher concentrations. *Org. Lett.* **11**, 3100–3103 (2009).
33. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851–1858 (2008).
34. Weber, K. P. *et al.* Whole genome sequencing highlights genetic changes associated with laboratory domestication of *C.elegans*. *PLoS ONE* **5**, e13922 (2010).
35. Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
36. Chalasani, S. H. *et al.* Dissecting a circuit for olfactory behaviour in *Caenorhabditis elegans*. *Nature* **450**, 63–70 (2007).

Sequence-based characterization of structural variation in the mouse genome

Binnaz Yalcin^{1*}, Kim Wong^{2*}, Avigail Agam^{1,3*}, Martin Goodson^{1*}, Thomas M. Keane², Xiangchao Gan¹, Christoffer Nellåker³, Leo Goodstadt¹, Jérôme Nicod¹, Amarjit Bhomra¹, Polinka Hernandez-Pliego¹, Helen Whitley¹, James Cleak¹, Rebekah Dutton¹, Deborah Janowitz^{1,4}, Richard Mott¹, David J. Adams² & Jonathan Flint¹

Structural variation is widespread in mammalian genomes^{1,2} and is an important cause of disease³, but just how abundant and important structural variants (SVs) are in shaping phenotypic variation remains unclear^{4,5}. Without knowing how many SVs there are, and how they arise, it is difficult to discover what they do. Combining experimental with automated analyses, we identified 711,920 SVs at 281,243 sites in the genomes of thirteen classical and four wild-derived inbred mouse strains. The majority of SVs are less than 1 kilobase in size and 98% are deletions or insertions. The breakpoints of 160,000 SVs were mapped to base pair resolution, allowing us to infer that insertion of retrotransposons causes more than half of SVs. Yet, despite their prevalence, SVs are less likely than other sequence variants to cause gene expression or quantitative phenotypic variation. We identified 24 SVs that disrupt coding exons, acting as rare variants of large effect on gene function. One-third of the genes so affected have immunological functions.

The pre-eminent organism for modelling the relationship between phenotype and genotype, including SVs, is the mouse, but our catalogue of SVs in this animal is incomplete⁶ and most of what we know about the impact of SVs on phenotypes comes from analyses of gene expression^{7,8}. Up to 28% of the between-strain variation in gene expression in haematopoietic stem and progenitor cells has been attributed to SVs⁷; SVs may account for between 66% and 74% of between-strain expression variation in kidney, liver, lung and testis⁸. Because gene expression variation is believed to contribute to variation in phenotypes in the whole organism⁹, SVs may turn out to have a major role in the genetic determination of many aspects of mouse biology.

Combining short-read paired-end mapping with experimental analyses (Supplementary Methods), we found SVs greater than 100 base pairs (bp) at 281,243 sites in the mouse genome, amounting to 711,920 SVs in thirteen classical and four wild-derived inbred strains of mice (Supplementary Table 1a), affecting 1.2% (33.0 Mb) and 3.7% (98.2 Mb) of the genome respectively (Supplementary Table 1b). Deletions, a category we can measure accurately, have a median size of 349 bp with modes at 100 bp and 6,400 bp (Supplementary Fig. 1a).

Our catalogue contains far more SVs than previously identified: 99.4% of SVs are simple and 0.6% are complex (Supplementary Table 1a), where simple SVs include insertions, deletions, inversions and copy number gains, and complex SVs consist of a mixture of events that abut each other. From experimental analyses of simple deletion SVs, we estimated an average false-negative rate of 17% in the classical inbred strains (Supplementary Tables 2a, b and 3a) and 24% in the wild-derived strains (Supplementary Table 2b); false-positive rates were below 5% for all strains (Supplementary Table 2c). False-negative rates for non-deletion simple SVs as well as complex SVs were higher than for simple deletions, ranging from 24% to 31% and 35% to 54% per strain, respectively (Supplementary Table 3b).

It proved difficult to obtain robust estimates of SVs smaller than 100 bp. Our best estimate of the rate of SVs between 30 and 100 bp is based on combining manual and automated methods over a region of 7.2 Mb (Supplementary Methods). Assuming that this region is typical, the rest of the genome (in classical laboratory strains) should contain approximately 49,000 SVs in this size range.

Microhomology at SV breakpoints, as well as the sequence content within SVs and the SV's ancestral state, were used to infer the likely mechanism of formation for simple SVs. To obtain breakpoint sequence, we performed *de novo* local assembly for 80.3% of deletions. Comparison of 1,314 predicted deletion breakpoints to the breakpoint delineated by PCR and sequencing (Supplementary Table 4) revealed that 57.7% of breakpoint predictions are exact and 86.5% are within 20 bp (Supplementary Table 5a). In cases where the local assembly strategy failed, we relied on the original breakpoint estimates obtained from the mapping of reads to the reference genome: 83.3% of these estimates are within 100 bp of the actual breakpoint (Supplementary Table 5b). Breakpoint accuracy for insertions, inversions and copy number gains is presented in Supplementary Table 5c, d and e, respectively.

Genome-wide estimates of the contribution of each mechanism to SV formation were derived from analysis of breakpoint sequence of deletions relative to C57BL/6J. We have highly accurate breakpoint sequence for this SV category, which should be unbiased with respect to ancestry. Using rat as an outgroup, we classified 19% of relative deletion SVs as ancestral deletions, 57% as ancestral insertions and the remainder (24%) were indeterminate (Supplementary Fig. 2).

SVs are most often due to retrotransposons (long interspersed nuclear elements (LINEs; 25%), long terminal repeats (LTRs; 14%) and short interspersed nuclear elements (SINEs; 15%)), followed by variable number tandem repeats (VNTRs) (15%) and pseudogenes (2%). Other mechanisms, not involving retrotransposons, account for 29% of SVs. Outgroup analysis showed that the transposon-associated SVs arose almost exclusively from ancestral insertion events (98.8%). Target site duplications (12–16 bp) surround the breakpoints of LINE and SINE derived SVs; shorter (6–8 bp) sequences are associated with LTR SVs (Supplementary Fig. 1b). Non-repeat-mediated SVs are mainly a result of ancestral deletion events (79%), and are associated with microhomologies up to 7 bp in length (Supplementary Fig. 1b), consistent with either microhomology-mediated break-induced replication¹⁰ or microhomology-mediated end joining¹¹.

Given their potential role in human disease¹², we were interested to document the occurrence of SVs that arise at the same genomic locus independently in unrelated strains (recurrent SVs). Non-allelic homologous recombination (NAHR) is the major mechanism for recurrent SVs¹³, whereas fork stalling and template switching and/or microhomology-mediated break-induced replication mechanisms may be important for non-recurrent SVs¹⁴.

¹The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. ²The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1HH, UK. ³MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3QX, UK. ⁴Department of Psychiatry and Psychotherapy, Ernst-Moritz-Arndt-Universität Greifswald Klinikum der Hansestadt Stralsund, Rostocker Chaussee 70, 18437 Stralsund, Germany.

*These authors contributed equally to this work.

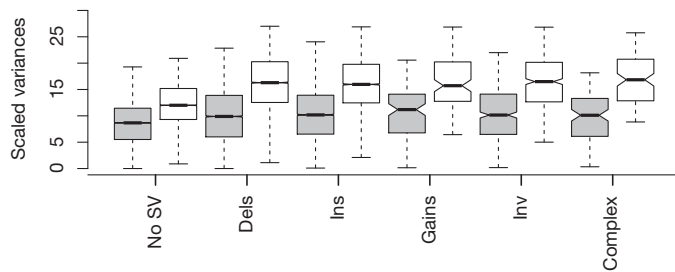


Figure 1 | Impact of SVs on gene expression. Within-strain (grey boxes) and between-strain (white boxes) gene expression variances for transcripts which are not overlapped by any structural variant (No SV) and for those which are overlapped. Within-strain variance is due to environmental effects; between-strain variance is due to environmental and genetic effects. The difference between the two variances is a measure of heritability. Six categories are shown: No SV, deletions (Dels), insertions (Ins), copy number gains (Gains), inversions (Inv), and complex rearrangements (Complex).

Using the SV breakpoints obtained from PCR sequencing (249 SV sites in eight strains, accounting for over 4,000 breakpoints, Supplementary Table 4), we identified SVs occurring at the same locus in different strains, but with different breakpoints, indicating independent origins. In the classical strains, only 2.5% of deletions at the same locus had different breakpoint sequences. However within all 17 strains we found multiple alleles at 12% of SVs, due almost entirely to the presence of different alleles originating from the wild-derived inbred strains. Consistent with the low frequency of recurrent SVs, breakpoint features associated with NAHR are rare. We estimated that 0.13% of deletions are due to NAHR, when we required a signature of ≥ 200 bp of $\geq 90\%$ sequence identity.

We assessed the impact of SVs on phenotypes by first estimating the proportion of heritability attributable to SVs⁸ from brain RNA-seq and found that no category accounts for more than 10% (Fig. 1). To determine if these results were specific to brain tissue, we analysed gene expression data for the eight founder strains of the heterogeneous stock (HS) population ($n = 5$ for each) from liver, measured on Illumina gene-expression arrays¹⁵. Mean heritability attributable to an SV, for transcripts overlapping one or more SVs, was 9.5%. Because many transcripts overlap multiple small SVs (median of 3, maximum of 216), we proposed that SV heritability might be related to the amount of gene overlapped. For each transcript we summed the amount of DNA overlapping a gene and expressed this as a proportion of the total length of the gene. SVs that overlap 50% or more of a gene make a large contribution to heritability: in brain tissue, such SVs contribute to 25% of the variance, compared to 7.8% for transcripts where SVs overlap less than 50% of the gene. However, large overlaps (50% or more) are rare, affecting less than 3% of transcripts. Thus, whereas SVs make a modest contribution to the overall heritability of expression variance, at individual transcripts they may be the main cause of between-strain differences in expression.

As another method to assess the impact of SVs on phenotype, we applied a test of functionality¹⁶ to 281,246 SVs in association with 100 phenotypes measured in over 2,000 HS mice¹⁷. We identified 290 quantitative trait loci (QTLs) where SVs were among the variants most likely to be functional, but in all these cases the SVs were only a subset of the total number of functional variants. We found a small but highly significant deficit in SVs among the functional variants (0.36% compared to 0.54% among the non-functional, $P < 10^{-16}$, $\chi^2 = 72.1$).

Whereas SVs make a relatively small contribution to the total amount of quantitative phenotypic variation, at a small number of QTLs they are the cause of variation. As shown in an accompanying paper¹⁸, larger effect QTLs are more likely to arise from SVs. We identified 12 QTLs where the SV overlapped a gene or flanking region (2 kb up and downstream), and where the QTL effect size is in the top 5% of the distribution. Table 1 lists these SVs, the genes they affect and the putative phenotype with which they are associated. Two associations have been directly tested: complementation of the deletion of the *H2-Ea* promoter has confirmed the effect of this SV on the T-cell phenotype¹⁹; analysis of a knock out of *Eps15* showed the predicted lower locomotor activity (Fig. 2a).

There are relatively few examples where an SV can be said unequivocally to delete one, or more, coding exons. Without nucleotide resolution accuracy we cannot be certain whether the breakpoint of an SV lies within an exon. Therefore to find SVs overlapping exons we used our most accurate and complete category of SV calls: deletions relative to C57BL/6J. We identified 210 that overlap exons (Ensembl Build 58); after removing pseudogenes, and genes not annotated as 'protein coding', we were left with 24 SVs that affect coding exons, including six that encompass a gene in its entirety (Table 2).

Five of the 24 SVs are already known^{20–24}, the remaining 19 are novel. A third of the genes affected are involved in immunity and infection. Our data expand current knowledge of the molecular architecture of these SVs. Figure 2b shows that antiviral genes *Trim5* and *Trim12a* are unique to C57BL/6J, due to segmental duplication²⁵. All the other strains contain only the *Trim12c* gene. Therefore the mouse contains a unique homologue of the human *TRIM5* gene. A similar analysis revealed that documented exonic changes in the defensin beta 8 gene (*Defb8*)²⁶ are linked to a previously undetected 3,192-bp ancestral insertion plus a 54-bp deletion (Table 2).

Our results are important in three respects. First, we find an unexpectedly large number of SVs with diverse molecular architecture, thus providing a catalogue of the most dynamic and variable regions of the mouse genome. Second, we were able to map almost 60% of deletions to base-pair resolution, allowing us to classify SVs by the mechanism that created them. In contrast to human SV studies, the great majority of SVs that we have discovered are non-recurrent rearrangements, based on two observations: among the classical strains, only 2.5% of deletions at the same locus had different breakpoint sequences and less than 1% of deletions are due to NAHR¹². Third, SVs have relatively little impact on gene function, a conclusion based on the following observations. We found that SVs overlapping a gene account for less than 10% of variation in gene

Table 1 | QTLs associated with SVs

Phenotype	Chromosome	SV start	SV stop	Ancestral event	Gene	SV overlap
Mean platelet volume	1	175158884*	175158885*	Ins (large)	<i>Fcer1a</i>	Upstream
OFT total activity	2	144402760	144402971	SINE Ins	<i>Sec23b</i>	Intron
Hippocampus cellular proliferation marker	4	49690362	49690363	Del (137 bp)	<i>Grin3a</i>	Intron
Home cage activity	4	108951263	108951264	IAP Ins (~6,400 bp)	<i>Eps15</i>	Upstream
T-cells: %CD3	4	130038388	130038389	SINE Ins (202 bp)	<i>Snrrp40</i>	Intron
Wound healing	7	90731819	90731820	IAP Ins (~6,400 bp)	<i>Tmc3</i>	Upstream
Red cells: mean cellular haemoglobin	7	111397607	111479433	Ins	<i>Trim5</i>	Exon
Red cells: mean cellular haemoglobin	7	111504989	111505193	Del	<i>Trim30b</i>	UTR
Red cells: mean cellular volume	8	87957244	87957245	LINE Ins (~500 bp)	<i>4921524J17Rik</i>	Upstream
Serum urea concentration	11	115106127	115106250	Del	<i>Tmem104</i>	UTR
Hippocampus cellular proliferation marker	13	113783196	113783359	Del	<i>Gm6320</i>	Upstream
T-cells: CD4/CD8 ratio	17	34483681	34483682	Del (629 bp)	<i>H2-Ea</i>	Upstream

Start and stop coordinates are given for MGSCv37 of the mouse reference genome. Unless there is an asterisk, coordinates refer to the exact coordinates as delineated by Sanger sequencing. IAP, intracisternal A particle; Ins, insertion; Del, deletion; LINE, long interspersed nuclear elements; SINE, short interspersed nuclear elements. OFT, open-field test. UTR, untranslated region.

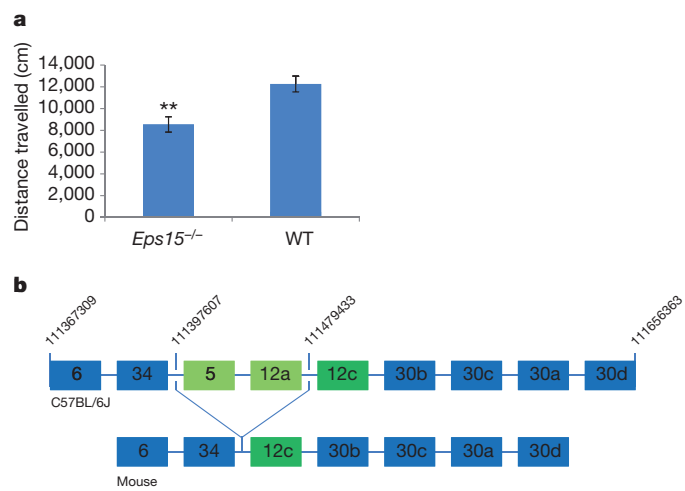


Figure 2 | Experimental analysis of SVs. **a**, Locomotor activity in *Eps15*^{-/-} mice. Activity was recorded during a period of 10 min in an open field arena. *n* = 7 for *Eps15*^{-/-} male mice and *n* = 16 for matched control wild type. ***P* value < 0.05. **b**, Schematic representation of the *Trim6-Trim30* genes cluster on chromosome 7. Boxes represent the sequential positions of the *Trim6*, *Trim34*, *Trim5/12* and *Trim30* genes. *Trim5* and *Trim12a* genes, which are only present in the C57BL/6J genome, occurred by segmental duplication of the *Trim12c* gene present in all 17 strains. The flanking *Trim34* and *Trim30* genes do not vary between strains. Coordinates are given for MGSCv37 assembly of the mouse reference genome.

expression, three to four times less than that found by studies using expression arrays^{7,8}. SVs overlapping exons are rare: because the frequency of insertions is equal to that of deletions, and because these two categories make up 98% of all SVs, extrapolating from the 24 SVs that delete exons, we predict that there are only about 50 SVs that directly overlap exons, or about 0.2% of the total burden of SVs in the genome. Finally, our analysis of the phenotypic consequences of SVs on QTLs for multiple phenotypes points to a relative deficit of SVs as the molecular basis of complex phenotypes. For the classical laboratory strains, single nucleotide polymorphisms (SNPs) and indels affect 0.5% of the genome, whereas on average 33 Mb (1.2%) of each classical laboratory strain falls into structurally variant regions of the genome. This implies that SVs are at least twice as likely to have phenotypic consequences than the combined effect of SNPs and indels. Yet we find that SVs contribute only 10%

to the heritability of gene expression, not the 50% implied by the genomic size argument.

It is important to note that conclusions based on our analysis of the HS outbred population may not apply to other outbred populations. The mouse population we tested is derived from inbred progenitors whose homozygosity will have purged their genomes of variants that could otherwise be maintained in heterozygous freely mating populations. Nevertheless, despite their relative rarity in the mouse genome, SVs that cause phenotype change are likely to provide biological insights out of proportion to their relative small contribution to phenotypic variance. We expect that the alleles we have described will provide a starting point for investigating the relationship between phenotype and genotype in mice.

METHODS SUMMARY

SV discovery. We used a combination of four computational methods: split-read mapping²⁷, mate-pair analysis²⁸, single-end cluster analysis (SECluster and RetroSeq), and read-depth²⁹. These methods identify deletions, insertions, inversions and copy number gains. We also derived methods to recognize other types of rearrangements, such as inversion plus insertion or inversion plus deletion, newly revealed from our experimental analysis.

Experimental analysis. We visually inspected short-read sequencing data using LookSeq³⁰ and manually detected SVs across mouse chromosome 19 in its entirety and a random set of other chromosomal regions. We analysed molecular structures of these SVs at nucleotide-level resolution using PCR and Sanger-based sequencing.

Outgroup analysis. The rat was used as an outgroup species to classify each mouse SV as either an ancestral deletion or an ancestral insertion. We predicted the ancestral state in the rat by estimating the size of the region in the rat genome that was homologous to the region that encompassed the mouse SV.

SV classification. We developed a machine learning method to classify SVs. The method used a random forest classifier, trained using sequence features within the SVs. Microhomology between breakpoints was determined by recording the longest sequence of bases that was identical between each breakpoint of each SV.

Functional impact of SVs. We tested whether an SV is likely to be functional using merge analysis¹⁶. The variances of expression data were calculated using ANOVA in the statistical software R using formulae described in ref. 8 and also by comparing a model where the expression value is explained by the strain, to a model in which the expression is explained by strain and whether or not the animal has an SV.

Received 5 July; accepted 4 August 2011.

1. Mills, R. E. *et al.* Mapping copy number variation by population-scale genome sequencing. *Nature* **470**, 59–65 (2011).

Table 2 | SVs affecting coding regions

Gene	Chromosome	SV start	SV stop	Ancestral event	Known function
<i>Soat1</i>	1	158394620	158401436	Del	Hair morphogenesis
<i>Olf1055</i>	2	86179898	86186982	IAP Ins	Olfactory
<i>Fcrl5</i>	3	87245084	87245947	Del	Infection and immunity
<i>Nes</i>	3	87780530	87780662	VNTR	Brain development
<i>Pglyrp3</i>	3	91831862	91835385	Del	Infection and immunity
<i>Skint4,3,9⁺</i>	4	111731004*	112272814*	Ins	Infection and immunity
<i>Fv1</i>	4	147244398	147245739	Del	Infection and immunity
<i>Ugt2b38</i>	5	87850554	87854999	Del	Metabolism
<i>Klrb1a</i>	6	128559593	128559740	Del	Infection and immunity
<i>Klri2</i>	6	129689526	129691211	Del	Infection and immunity
<i>Tas2r120⁺</i>	6	132580541	132613777	Del+Ins 326-bp	Taste
<i>Tas2r103</i>	6	132985563	132986696	Del	Taste
<i>Zfp607⁺</i>	7	28646761	28671650	Del	DNA-binding
<i>Krtap5-5</i>	7	149415121	149415210	VNTR	Hair formation
<i>Trim5,12a⁺</i>	7	111397607	111479433	Ins	Infection and immunity
<i>Defb8</i>	8	19447465	19450575	Ins+54-bp Del	Infection and immunity
<i>Zfp872</i>	9	22004856	22005023	VNTR	DNA-binding
<i>Olf913</i>	9	38402589	38403498	Del	Olfactory
<i>Rtp3</i>	9	110889280	110889465	VNTR	Bone density
<i>Nlrp1c⁺</i>	11	71046193*	71101410*	Ins	Embryonic development
<i>Fam110c</i>	12	31759321	31759461	VNTR	Cell migration
<i>Olf234</i>	15	98328544	98328861	Del	Olfactory
<i>Krtap16-1</i>	16	88874294	88874392	VNTR	Hair formation
<i>Amd2⁺</i>	18	64607747	64609669	Ins	Biosynthesis of polyamines

Start and stop coordinates are given for MGSCv37 assembly of the mouse reference genome. Plus signs (+) indicate that the structural variant overlaps the entire gene. Unless there is an asterisk, coordinates refer to the exact coordinates as delineated by Sanger sequencing. VNTR, variable number tandem repeat.

2. Quinlan, A. R. *et al.* Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. *Genome Res.* **20**, 623–635 (2010).
3. Zhang, F., Gu, W., Hurler, M. E. & Lupski, J. R. Copy number variation in human health, disease, and evolution. *Annu. Rev. Genomics Hum. Genet.* **10**, 451–481 (2009).
4. Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2010).
5. Stranger, B. E. *et al.* Population genomics of human gene expression. *Nature Genet.* **39**, 1217–1224 (2007).
6. Agam, A. *et al.* Elusive copy number variation in the mouse genome. *PLoS ONE* **5**, e12839 (2010).
7. Cahan, P., Li, Y., Izumi, M. & Graubert, T. A. The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells. *Nature Genet.* **41**, 430–437 (2009).
8. Henriksen, C. N. *et al.* Segmental copy number variation shapes tissue transcriptomes. *Nature Genet.* **41**, 424–429 (2009).
9. Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genet.* **37**, 710–717 (2005).
10. Zhang, F. *et al.* The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nature Genet.* **41**, 849–853 (2009).
11. Ma, J. L., Kim, E. M., Haber, J. E. & Lee, S. E. Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol. Cell. Biol.* **23**, 8820–8828 (2003).
12. Stankiewicz, P. & Lupski, J. R. Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* **61**, 437–455 (2010).
13. Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82 (2002).
14. Hastings, P. J., Ira, G. & Lupski, J. R. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* **5**, e1000327 (2009).
15. Huang, G. J. *et al.* High resolution mapping of expression QTLs in heterogeneous stock mice in multiple tissues. *Genome Res.* **19**, 1133–1140 (2009).
16. Yalcin, B., Flint, J. & Mott, R. Using progenitor strain information to identify quantitative trait nucleotides in outbred mice. *Genetics* **171**, 673–681 (2005).
17. Valdar, W. *et al.* Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nature Genet.* **38**, 879–887 (2006).
18. Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* doi:10.1038/nature10413 (this issue).
19. Yalcin, B. *et al.* Commercially available outbred mice for genome-wide association studies. *PLoS Genet.* **6**, e1001085 (2010).
20. Best, S., Le Tissier, P., Towers, G. & Stoye, J. P. Positional cloning of the mouse retrovirus restriction gene *Fv1*. *Nature* **382**, 826–829 (1996).
21. Boyden, L. M. *et al.* *Skint1*, the prototype of a newly identified immunoglobulin superfamily gene cluster, positively selects epidermal $\gamma\delta$ T cells. *Nature Genet.* **40**, 656–662 (2008).
22. Nelson, T. M., Munger, S. D. & Boughter, J. D. Jr. Haplotypes at the *Tas2r* locus on distal chromosome 6 vary with quinine taste sensitivity in inbred mice. *BMC Genet.* **6**, 32 (2005).
23. Persson, K., Heby, O. & Berger, F. G. The functional intronless S-adenosylmethionine decarboxylase gene of the mouse (*Amd-2*) is linked to the ornithine decarboxylase gene (*Odc*) on chromosome 12 and is present in distantly related species of the genus *Mus*. *Mamm. Genome* **10**, 784–788 (1999).
24. Wu, B. *et al.* Mutations in sterol O-acyltransferase 1 (*Soat1*) result in hair interior defects in AKR/J mice. *J. Invest. Dermatol.* **130**, 2666–2668 (2010).
25. Tareen, S. U., Sawyer, S. L., Malik, H. S. & Emerman, M. An expanded clade of rodent *Trim5* genes. *Virology* **385**, 473–483 (2009).
26. Taylor, K. *et al.* Defensin-related peptide 1 (*Defr1*) is allelic to Defb8 and chemoattracts immature DC and CD4+ T cells independently of CCR6. *Eur. J. Immunol.* **39**, 1353–1360 (2009).
27. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
28. Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature Methods* **6**, 677–681 (2009).
29. Simpson, J. T., McIntyre, R. E., Adams, D. J. & Durbin, R. Copy number variant detection in inbred strains from short read sequence data. *Bioinformatics* **26**, 565–567 (2010).
30. Mankske, H. M. & Kwiatkowski, D. P. LookSeq: a browser-based viewer for deep sequencing data. *Genome Res.* **19**, 2125–2132 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Whitley, G. Durrant, A. M. Hammond, D. J. Fabrigar, L. Chen, M. Johannesson, E. Cong and G. Blázquez for helping B.Y. with various laboratory-based work. We also thank C. P. Ponting for comments on the manuscript. This project was supported by The Medical Research Council, UK, and the Wellcome Trust. D.J.A. is supported by Cancer Research UK.

Author Contributions D.J.A. and J.F. conceived the study and directed the research. J.F. wrote the core of the paper. K.W. and T.K. performed the genome-wide SV discovery and local assembly for SV breakpoint resolution. K.W. carried out the sensitivity and specificity analyses. K.W. and B.Y. liaised regularly to integrate experimental work into genome-wide SV discovery pipeline. This resulted in a highly accurate map of SV across the mouse genome, essential to downstream analyses. A.B., P.H.P., H.W., J.C., R.D. and D.J. carried out experimental work, led by B.Y. A.B. and B.Y. analysed Sanger-based sequencing data, resolved SV breakpoints at nucleotide-level resolution and inferred mechanism of SV formation. M.G. performed the genome-wide SV mechanism of formation and outgroup analysis, with contributions from A.A. and B.Y.; J.F. and A.A. analysed functional impact of SVs on expression and phenotypes. C.N., L.G., J.N., A.A. and R.M. carried out additional analyses. B.Y. characterized function of individual SV examples.

Author Information Data sets described here will be available under study accession number estd118 from the Database of Genomic Variants archive (DGVa) at <http://www.ebi.ac.uk/dgva/page.php>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.F. (jf@well.ox.ac.uk) or D.J.A. (da1@sanger.ac.uk).

FADD prevents RIP3-mediated epithelial cell necrosis and chronic intestinal inflammation

Patrick-Simon Welz¹, Andy Wullaert¹, Katerina Vlantis¹, Vangelis Kondylis¹, Vanesa Fernández-Majada¹, Maria Ermolaeva¹, Petra Kirsch², Anja Sterner-Kock³, Geert van Loo⁴ & Manolis Pasparakis¹

Intestinal immune homeostasis depends on a tightly regulated cross talk between commensal bacteria, mucosal immune cells and intestinal epithelial cells (IECs)^{1–4}. Epithelial barrier disruption is considered to be a potential cause of inflammatory bowel disease; however, the mechanisms regulating intestinal epithelial integrity are poorly understood^{1,5}. Here we show that mice with IEC-specific knockout of FADD (FADD^{IEC-KO}), an adaptor protein required for death-receptor-induced apoptosis⁶, spontaneously developed epithelial cell necrosis, loss of Paneth cells, enteritis and severe erosive colitis. Genetic deficiency in RIP3, a critical regulator of programmed necrosis^{7–9}, prevented the development of spontaneous pathology in both the small intestine and colon of FADD^{IEC-KO} mice, demonstrating that intestinal inflammation is triggered by RIP3-dependent death of FADD-deficient IECs. Epithelial-specific inhibition of CYLD, a deubiquitinase that regulates cellular necrosis¹⁰, prevented colitis development in FADD^{IEC-KO} but not in NEMO^{IEC-KO} mice¹¹, showing that different mechanisms mediated death of colonic epithelial cells in these two models. In FADD^{IEC-KO} mice, TNF deficiency ameliorated colon inflammation, whereas MYD88 deficiency and also elimination of the microbiota prevented colon inflammation, indicating that bacteria-mediated Toll-like-receptor signalling drives colitis by inducing the expression of TNF and other cytokines. However, neither CYLD, TNF or MYD88 deficiency nor elimination of the microbiota could prevent Paneth cell loss and enteritis in FADD^{IEC-KO} mice, showing that different mechanisms drive RIP3-dependent necrosis of FADD-deficient IECs in the small and large bowel. Therefore, by inhibiting RIP3-mediated IEC necrosis, FADD preserves epithelial barrier integrity and antibacterial defence, maintains homeostasis and prevents chronic intestinal inflammation. Collectively, these results show that mechanisms preventing RIP3-mediated epithelial cell death are critical for the maintenance of intestinal homeostasis and indicate that programmed necrosis of IECs might be implicated in the pathogenesis of inflammatory bowel disease, in which Paneth cell and barrier defects are thought to contribute to intestinal inflammation.

To study the role of FADD in the intestinal epithelium we crossed mice carrying loxP-flanked *Fadd* alleles (FADD^{FL}) with villin-Cre transgenics (Fig. 1a, b). FADD^{IEC-KO} mice were born normally but developed a spontaneous phenotype resulting in the death of about 50% of these animals before weaning. Surviving FADD^{IEC-KO} mice showed reduced body weight and diarrhoea, indicating that they suffered from intestinal disease. High-resolution mini-endoscopy revealed mucosal thickening, ulceration and altered vascularisation in the colon of FADD^{IEC-KO} mice (Fig. 1c). Macroscopically, colons from FADD^{IEC-KO} mice were shorter and thicker compared to controls (Fig. 1d). Histological analysis of colon sections from 10-week-old FADD^{IEC-KO} mice revealed severe transmural inflammation affecting the entire colon with large areas of

epithelial erosion accompanied by crypt abscesses (Fig. 1e, f). Whereas FADD^{IEC-KO} mice developed spontaneous colitis with 100% penetrance, none of their FADD^{FL} littermates housed in the same cages showed any signs of colon inflammation, showing that colitis development was determined by FADD deficiency in IECs and was not transferable horizontally to co-housed wild-type littermates. Dying epithelial cells and early signs of inflammation were detectable in 2-week-old FADD^{IEC-KO} mice, and crypt abscesses together with increased immune cell infiltration and epithelial hyperproliferation were observed in 3-week-old animals (Fig. 1e). Increased cytokine and chemokine expression (Fig. 1g) and infiltration of F4/80⁺ and Gr-1⁺ myeloid cells initially, but also T and B lymphocytes in older animals, were detected in the colons of FADD^{IEC-KO} mice (Supplementary Fig. 1). However, FADD^{IEC-KO}/Rag1^{-/-} mice developed colitis, showing that T and B cells are not essential for colon inflammation in this model (data not shown). Thus, epithelial-specific FADD ablation caused the spontaneous development of severe colon inflammation driven primarily by an innate immune response. Immunostaining for Ki67 and cyclin D1 mRNA expression analysis revealed ongoing epithelial regeneration with increased epithelial cell proliferation in the colon of FADD^{IEC-KO} mice (Supplementary Fig. 2). In some cases, dysplastic crypts were detected in colons from 10-week-old FADD^{IEC-KO} mice, indicating that the chronic inflammatory and regenerative lesions occasionally resulted in epithelial dysplasia (data not shown).

Histological analysis of colon sections from FADD^{IEC-KO} mice revealed increased death of crypt epithelial cells in 2- to 3-week-old mice (Fig. 1e), indicating that epithelial cell death occurs early on during lesion development. Consistent with the well-established role of FADD as a mediator of apoptosis^{6,12,13}, many of the dead cells observed in crypt abscesses and early dying cells in the crypt epithelium did not stain with antibodies recognizing active caspase 3 (Fig. 2a), indicating that these cells did not die by apoptosis. Electron microscopy revealed epithelial cells showing signs of cellular necrosis such as disruption of the plasma membrane, swollen organelles, lack of chromatin condensation in the nucleus and a cytoplasm with lower electron density (Fig. 2b and Supplementary Fig. 3), indicating that FADD-deficient IECs mainly undergo necrotic cell death. Caspase inhibition and lack of FADD or caspase 8 were previously shown to sensitize certain cell types to a particular type of necrotic death, termed programmed necrosis or necroptosis, which is induced by death receptors such as TNFR1 and requires the kinases RIP1 and RIP3 (refs 7–9, 14–16). We therefore reasoned that programmed necrosis of FADD-deficient IECs could be a critical early event triggering colitis in FADD^{IEC-KO} mice. Primary IECs lacking FADD expressed increased levels of RIP3, an essential mediator of programmed necrosis^{7–9,14,17}, indicating that FADD deficiency might sensitize colonic epithelial cells to RIP3-dependent necrosis (Fig. 2c, d). To assess unambiguously the role of RIP3 in epithelial cell death

¹Institute for Genetics, Centre for Molecular Medicine (CMMC), and Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Zùlpicher Str. 47a, 50674 Cologne, Germany. ²Tierforschungszentrum, University of Ulm, Albert-Einstein-Allee 11, D-89081 Ulm, Germany. ³Center for Experimental Medicine, Uniklinik Köln, University of Cologne, Robert-Kochstr. 10, 50931 Cologne, Germany. ⁴Department for Molecular Biomedical Research, VIB, and Department of Biomedical Molecular Biology, Ghent University, Technologiepark 927, 9052 Ghent, Belgium.

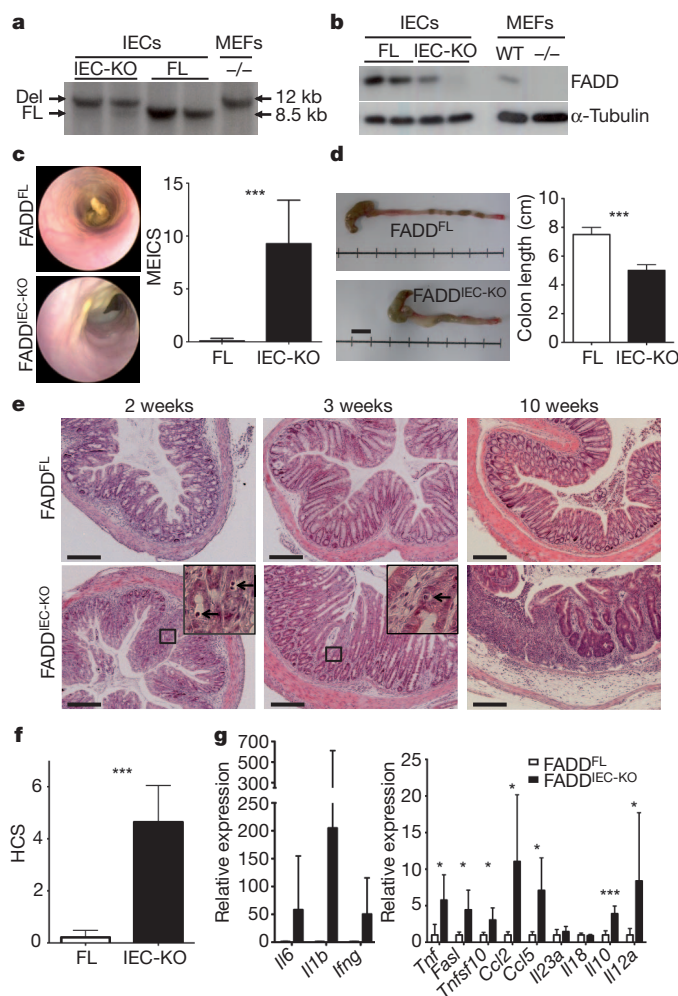


Figure 1 | Mice with IEC-specific ablation of FADD spontaneously develop severe colitis. **a**, Southern blot of EcoRV-digested genomic DNA (a) and immunoblot of protein extracts (b) from colonic IECs from FADD^{FL} and FADD^{IEC-KO} mice and from wild-type (WT) and *Fadd*^{-/-} (-/-) MEFs. Del, deleted; FL, loxP flanked. α-Tubulin serves as loading control. **c**, Representative endoscopic images and quantification of murine endoscopic index of colitis severity (MEICS) in FADD^{FL} (*n* = 31) and FADD^{IEC-KO} littermates (*n* = 37). **d**, Representative colon pictures and quantification of colon length in FADD^{IEC-KO} (*n* = 4) and FADD^{FL} littermates (*n* = 5). Scale bar, 1 cm. **e**, Representative histological images from haematoxylin & eosin stained colon sections from FADD^{FL} and FADD^{IEC-KO} mice. Arrows in insets indicate dying epithelial cells in 2- and 3-week-old mice. Scale bars, 100 μm. **f**, Histological colitis score (HCS) measuring severity of inflammation and tissue damage in colons from FADD^{FL} (*n* = 7) and FADD^{IEC-KO} (*n* = 7) mice. **g**, Quantitative polymerase chain reaction with reverse transcription (qRT-PCR) analysis of cytokine and chemokine expression in colons from 10-week-old FADD^{IEC-KO} and FADD^{FL} littermates (*n* = 5–8 for each genotype). All graphs show mean values ± standard deviation (s.d.). **P* ≤ 0.05; ****P* ≤ 0.005.

and colitis development, we crossed FADD^{IEC-KO} mice with RIP3-deficient mice¹⁸. FADD^{IEC-KO}/*Ripk3*^{-/-} mice developed normally and did not show macroscopic signs of disease and the early lethality associated with epithelial FADD deficiency. Moreover, colon sections from double-deficient FADD^{IEC-KO}/*Ripk3*^{-/-} mice showed a normal histology without signs of epithelial cell death or inflammation (Fig. 2e), demonstrating that RIP3 is essential for the spontaneous death of epithelial cells and the development of colitis in FADD^{IEC-KO} mice.

The deubiquitinating enzyme CYLD was identified as an important mediator of TNFR1-induced necrosis, presumably acting by deubiquitinating RIP1 to facilitate the formation of the RIP1/RIP3-containing 'necrosome' complex^{10,14}. To assess whether CYLD catalytic activity

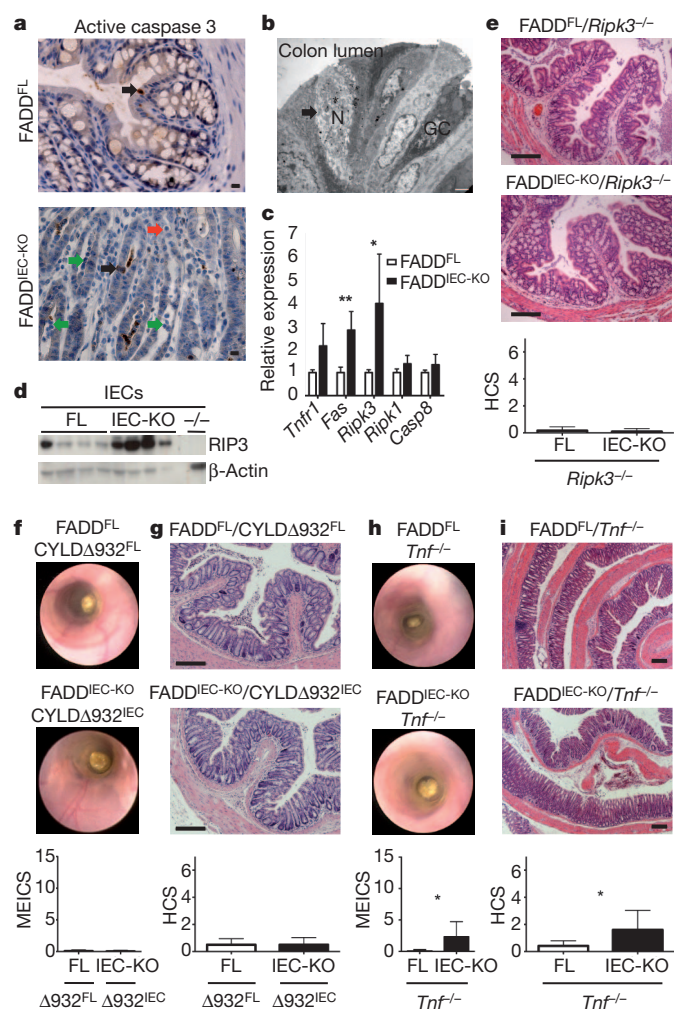


Figure 2 | RIP3- and CYLD-dependent necrosis of IECs triggers colitis in FADD^{IEC-KO} mice. **a**, Colon sections from FADD^{IEC-KO} and FADD^{FL} mice were immunostained for active caspase 3 (brown) and counterstained with haematoxylin (blue). Black arrow indicates apoptotic, red arrow shows early resident necrotic and green arrows show late detached necrotic epithelial cells. **b**, Representative electron microscopy picture showing a necrotic epithelial cell (arrow) in the proximal colon of a FADD^{IEC-KO} mouse. GC, goblet cell; N, nucleus. Scale bar, 2 μm. **c**, qRT-PCR showed increased *Tnfr1*, *Fas* and *Ripk3* messenger RNA expression in colonic IECs from FADD^{IEC-KO} (*n* = 3) compared to FADD^{FL} (*n* = 4) mice. **d**, Immunoblotting for RIP3 and β-actin (loading control) in colonic IECs from FADD^{IEC-KO} (IEC-KO), FADD^{FL} (FL) and *Ripk3*^{-/-} (-/-) mice. Lanes represent IECs from individual mice. **e**, Representative histological images and quantification of HCS in colon sections from FADD^{IEC-KO}/*Ripk3*^{-/-} (*n* = 5) and FADD^{FL}/*Ripk3*^{-/-} (*n* = 3) littermates. **f**, Representative endoscopic images and quantification of MEICS in colons from FADD^{IEC-KO}/*CYLD*Δ932^{IEC} (*n* = 11) and FADD^{FL}/*CYLD*Δ932^{FL} (*n* = 7) littermates. **g**, Representative histological images and quantification of HCS in colon sections from FADD^{IEC-KO}/*CYLD*Δ932^{IEC} (*n* = 10) and FADD^{FL}/*CYLD*Δ932^{FL} (*n* = 6) littermates. **h**, Representative endoscopic images and quantification of MEICS in colons from FADD^{IEC-KO}/*Tnfr1*^{-/-} (*n* = 6) and FADD^{FL}/*Tnfr1*^{-/-} (*n* = 6) littermates. **i**, Representative histological images and quantification of HCS in colon sections from FADD^{IEC-KO}/*Tnfr1*^{-/-} (*n* = 5) and FADD^{FL}/*Tnfr1*^{-/-} (*n* = 6) littermates. All graphs show mean values ± s.d. **P* ≤ 0.05; ***P* ≤ 0.01. Scale bars: **a**, 10 μm; **b**, 2 μm, **e** and **i**, 100 μm.

was required for spontaneous programmed necrosis of FADD-deficient IECs, we crossed FADD^{IEC-KO} mice with mice carrying conditional *CYLD*Δ932^{FL} alleles, which upon Cre recombination produce truncated *CYLD*Δ932 protein lacking the last 20 amino acids that are essential for its deubiquitinase activity¹⁹ (Supplementary Fig. 4a). Mouse embryonic fibroblasts (MEFs) homozygously expressing

truncated CYLDA932 were protected from TNF-induced death in the presence of the pan-caspase inhibitor zVAD-fmk, demonstrating that CYLD deubiquitinase activity is important for necroptosis (Supplementary Fig. 4d). FADD^{IEC-KO}/CYLDA932^{IEC} mice, which lack FADD and at the same time express catalytically inactive CYLDA932 specifically in IECs (Supplementary Fig. 4b, c), did not show any macroscopic signs of disease such as reduced weight or diarrhoea. In addition, endoscopic (Fig. 2f) and histological analysis (Fig. 2g) did not reveal signs of inflammation or epithelial destruction in the colons of FADD^{IEC-KO}/CYLDA932^{IEC} mice. Therefore, CYLD catalytic activity is required for colonocyte death and colitis development in FADD^{IEC-KO} mice. CYLD was also shown to contribute to TNF-induced apoptosis in the presence of Smac-mimetic compounds inducing the degradation of cIAP1/2 (ref. 20). We therefore investigated whether CYLD catalytic activity is also required for epithelial cell death and colitis development in NEMO^{IEC-KO} mice, which show increased IEC apoptosis and spontaneously develop severe chronic colon inflammation¹¹. In contrast to FADD^{IEC-KO}/CYLDA932^{IEC} mice, NEMO^{IEC-KO}/CYLDA932^{IEC} animals developed severe colitis similarly to single NEMO^{IEC-KO} mice as assessed by endoscopic and histological analysis (Supplementary Fig. 5). Therefore, inhibition of CYLD catalytic activity prevented epithelial cell death and colitis development in FADD^{IEC-KO} but not in NEMO^{IEC-KO} mice, indicating that the IECs in these two models die by different mechanisms.

TNF is a potent inducer of necroptosis and has an important pathogenic role in the development of intestinal inflammation in both humans and animal models^{21,22}. We therefore crossed FADD^{IEC-KO} mice with TNF-deficient mice to investigate whether TNF is implicated in colitis development in this model. Endoscopic and histological analysis of FADD^{IEC-KO}/Tnf^{-/-} mice showed areas of mild focal epithelial lesions in the colonic mucosa characterized by crypt elongation and the presence of inflammatory infiltrates (Fig. 2h, i). However, endoscopic and histological inflammation scores in FADD^{IEC-KO}/Tnf^{-/-} mice were significantly lower compared to FADD^{IEC-KO} animals, showing that TNF deficiency strongly ameliorated but could not completely prevent colon inflammation. Thus, TNF has an important role but TNF-independent mechanisms also contribute to the pathogenesis of colitis in FADD^{IEC-KO} mice.

Epithelial cell death could trigger colitis by disrupting the epithelial barrier thus allowing commensal bacteria to invade the mucosa, where they could induce inflammation by activating Toll-like-receptor (TLR) signalling on mucosal immune cells. To address the potential role of TLR signalling in colitis development, we crossed FADD^{IEC-KO} mice with mice lacking MYD88, an essential adaptor molecule for signalling downstream of most TLRs. FADD^{IEC-KO}/Myd88^{-/-} mice did not show macroscopic, endoscopic or histological signs of colon inflammation (Fig. 3a, b) demonstrating that MYD88-dependent signalling is essential for colitis development and suggesting that inflammation could be driven by commensal bacteria. Indeed, treatment with broad-spectrum antibiotics strongly attenuated colon inflammation in FADD^{IEC-KO} mice (Supplementary Fig. 6). Furthermore, FADD^{IEC-KO} mice raised in germ-free conditions did not show any endoscopic or histological signs of colon inflammation (Fig. 3c, d). When young adult germ-free FADD^{IEC-KO} mice were conventionalized by exposure to the microbiota of SPF mice they rapidly developed severe intestinal disease leading to the death of 4 out of 12 animals within 7 days (Supplementary Table 1). Endoscopic and histological analysis of colons from conventionalized FADD^{IEC-KO} mice 7 days after co-housing revealed severe colitis with mucosal thickening, epithelial erosion and transmural inflammation (Fig. 3c, d). Collectively, these results show that commensal bacteria induce colitis in FADD^{IEC-KO} mice by activating MYD88-dependent TLR signalling. Although the cellular targets of bacteria-induced TLR signalling in this model remain unclear at present, it is likely that the microbiota induces colitis development by activating the expression of TNF and other proinflammatory cytokines in mucosal immune cells. Indeed, conventionalization induced

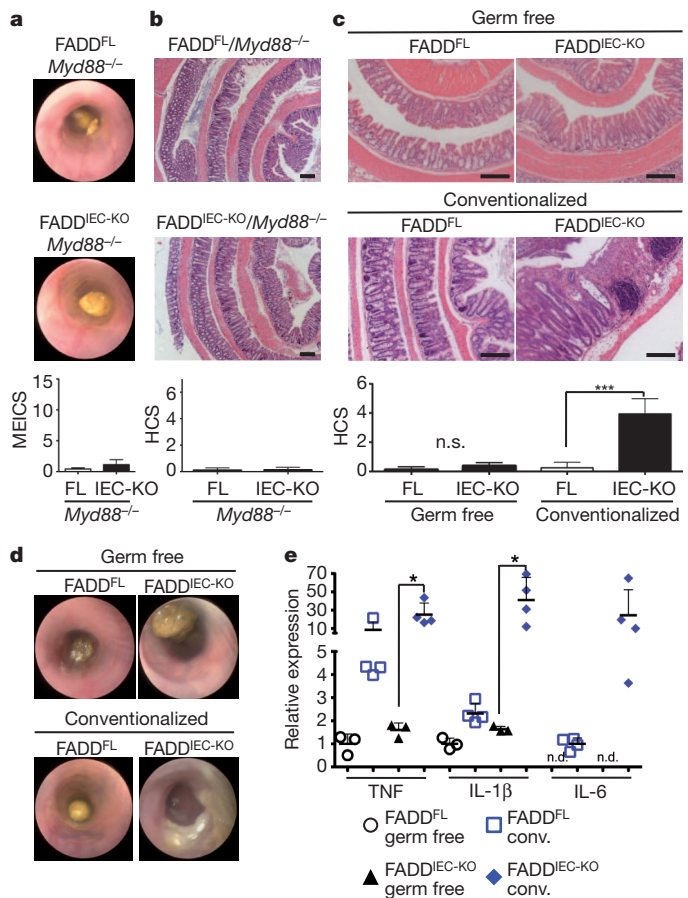


Figure 3 | Spontaneous colitis development in FADD^{IEC-KO} mice requires MYD88-dependent signalling and the presence of the microbiota.

a, Representative endoscopic images and quantification of MEICS in colons from FADD^{IEC-KO}/Myd88^{-/-} (n=9) and FADD^{FL}/Myd88^{-/-} (n=6) littermates. **b**, Representative histological images and quantification of HCS in colon sections from FADD^{IEC-KO}/Myd88^{-/-} (n=9) and FADD^{FL}/Myd88^{-/-} (n=6) littermates. **c**, Representative histological images and quantification of HCS in colon sections from germ-free FADD^{FL} (n=5), germ-free FADD^{IEC-KO} (n=7), conventionalized FADD^{FL} (n=8) and conventionalized FADD^{IEC-KO} (n=8) mice. n.s., not significant. **d**, Representative endoscopic colon images of germ-free and conventionalized FADD^{FL} and FADD^{IEC-KO} mice. **e**, qRT-PCR analysis of TNF, IL-1β and IL-6 expression in colons of conventionalized (conv.) FADD^{FL} and FADD^{IEC-KO} mice compared to germ-free animals. n.d., not detectable. Graphs show mean values ± s.d. *P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.005. Scale bars, 100 μm.

increased expression of TNF, IL-1β and IL-6 in the colons of both control FADD^{FL} and FADD^{IEC-KO} mice (Fig. 3e), supporting the notion that bacteria trigger colitis by inducing cytokine expression in the colonic mucosa.

In addition to colitis, FADD^{IEC-KO} mice also developed enteritis characterized by altered intestinal architecture with blunted and fused villi, mucosal oedema and increased cellularity of the lamina propria (Fig. 4a). Consistent with inflammatory changes, increased numbers of granulocytes and increased epithelial cell proliferation were detected in the small intestine of FADD^{IEC-KO} mice (Supplementary Fig. 7). In addition, small intestinal crypts in FADD^{IEC-KO} mice contained strongly reduced numbers of Paneth cells, as identified by their characteristic morphology with a large cytoplasm filled with eosinophilic granules (Fig. 4a). Immunostaining for lysozyme, an early marker of Paneth cells, confirmed the strongly reduced Paneth cell numbers in FADD^{IEC-KO} mice (Fig. 4b). Paneth cells are believed to contribute to the intestinal antibacterial defence by releasing antimicrobial factors stored in cytoplasmic granules. Consistent with the reduced Paneth cell numbers, we detected impaired expression of antimicrobial factors

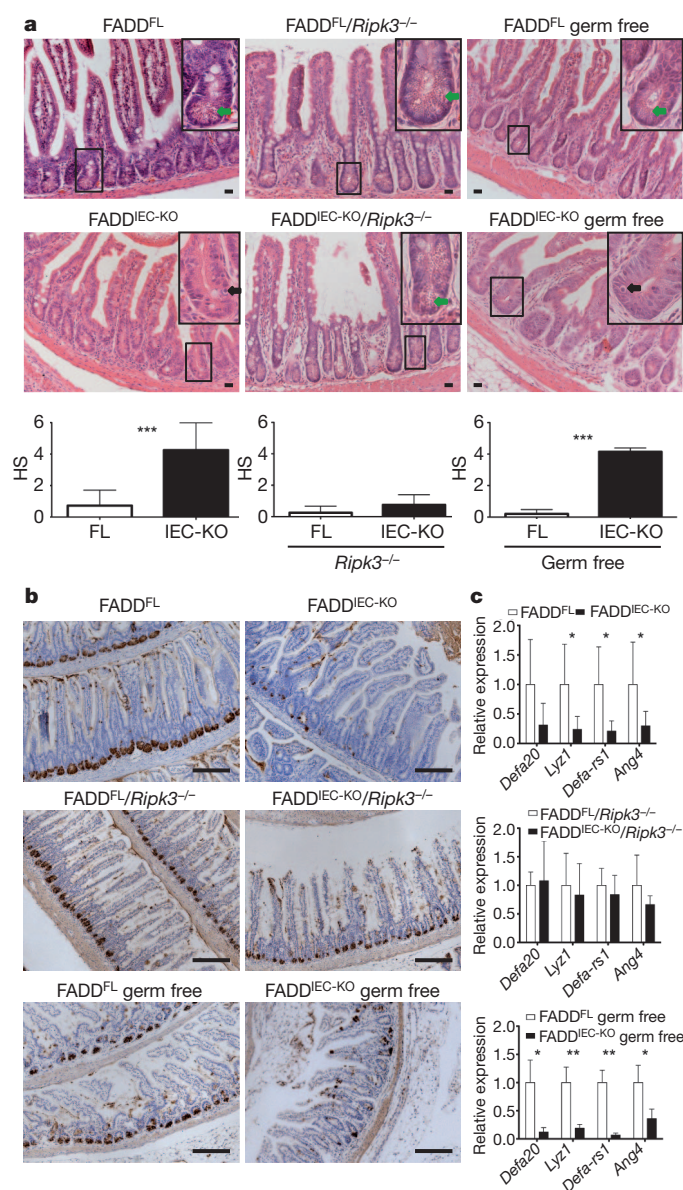


Figure 4 | Spontaneous development of enteritis and loss of Paneth cells in FADD^{IEC-KO} mice requires RIP3-mediated necrosis of IECs but does not depend on the microbiota. **a**, Representative histological images and quantification of histological score (HS) in small intestinal sections from mice with the indicated genotypes. FADD^{IEC-KO} mice show death of small intestinal IECs, enteritis and loss of Paneth cells. This pathology was prevented by RIP3 deficiency but persisted in germ-free FADD^{IEC-KO} mice. Green arrows indicate Paneth cells, black arrows indicate dying crypt epithelial cells in insets. FADD^{FL} (n = 11), FADD^{IEC-KO} (n = 8), FADD^{FL}/Ripk3^{-/-} (n = 6), FADD^{IEC-KO}/Ripk3^{-/-} (n = 4), germ-free FADD^{FL} (n = 5), germ-free FADD^{IEC-KO} (n = 7). **b**, Small intestinal sections were immunostained for lysozyme (brown) and counterstained with haematoxylin (blue). Paneth cell loss was prevented by RIP3 deficiency but persisted in germ-free FADD^{IEC-KO} mice. **c**, Expression of the Paneth-cell-specific genes *Defa20*, *Lyz1*, *Defa-rs1* and *Ang4* was measured by qRT-PCR in small intestinal mRNA samples from mice with the indicated genotypes. FADD^{FL} (n = 6), FADD^{IEC-KO} (n = 6), FADD^{FL}/Ripk3^{-/-} (n = 3), FADD^{IEC-KO}/Ripk3^{-/-} (n = 5), germ-free FADD^{FL} (n = 3), germ-free FADD^{IEC-KO} (n = 3). Graphs show mean values \pm s.d. **P* \leq 0.05; ***P* \leq 0.01. Scale bars: **a**, 10 μ m; **b**, 100 μ m.

including lysozyme (*Lyz1*), α -defensin 20 (*Defa20*), α -defensin-related sequence 1 (*Defa-rs1*) and angiogenin 4 (*Ang4*) in the ileum of FADD^{IEC-KO} mice (Fig. 4c). Increased numbers of dying epithelial cells that did not stain with antibodies recognizing active caspase 3 were detected in small intestinal crypts from FADD^{IEC-KO} mice

(Supplementary Fig. 8), suggesting that caspase-independent death of FADD-deficient IECs could contribute to the development of enteritis. Indeed, RIP3 deficiency prevented epithelial cell death, Paneth cell loss and enteritis in FADD^{IEC-KO} mice (Fig. 4 and Supplementary Fig. 8), demonstrating that, similarly to the colitis, the small intestinal lesions are caused by RIP3-dependent programmed necrosis of FADD-deficient IECs. However, in contrast to our findings in the colon, neither the absence of the microbiota nor genetic deficiency in MYD88, TNF or CYLD could prevent Paneth cell loss and enteritis in FADD^{IEC-KO} mice (Fig. 4 and Supplementary Figs 7–9). These results indicate that different, perhaps cell-intrinsic mechanisms induce RIP3-dependent programmed necrosis of small intestinal IECs. Interestingly, mice with epithelial-specific ablation of the transcription factor XBP1, a critical regulator of the endoplasmic reticulum stress response, developed spontaneous enteritis and Paneth cell loss²³, and mutations affecting autophagy also caused Paneth cell abnormalities²⁴. These findings indicated that owing to their highly secretory activity Paneth cells are particularly sensitive to endoplasmic reticulum stress and autophagy defects¹. Although the mechanisms inducing RIP3-dependent necrosis of small intestinal IECs in FADD^{IEC-KO} mice remain unclear at present, it is tempting to speculate that pathways linked to endoplasmic reticulum stress or autophagy might be implicated in triggering programmed necrosis of epithelial cells, Paneth cell loss and enteritis in these animals. Paneth cell loss might also be important for the development of colitis in FADD^{IEC-KO} mice, as reduced expression of Paneth-cell-derived antimicrobial factors could induce alterations in the microbiota, which might contribute to the bacteria-driven mechanisms triggering programmed necrosis of colonic epithelial cells and inflammation.

Our *in vivo* genetic mouse model studies indicate that mechanisms regulating programmed necrosis in IECs might be relevant for the pathogenesis of chronic intestinal inflammation in humans. Paneth cell defects leading to impaired antimicrobial peptide expression have been suggested to contribute to the pathogenesis of inflammatory bowel disease¹. Interestingly, epithelial patchy necrosis has been detected in the colon of Crohn's disease patients, indicating that necrotic death of IECs might be implicated in human colon inflammation²⁵. In this context, the potential capacity of enteropathogenic bacteria or viruses to induce TNF expression in the intestinal mucosa and at the same time to modulate epithelial responses to TNF signalling, for example by expressing inhibitors of apoptosis or programmed necrosis¹⁷, might be critical for triggering acute episodes of intestinal inflammation or precipitating chronic inflammatory bowel disease in genetically susceptible individuals. Moreover, our results indicate that anti-TNF therapy, shown to be highly effective in a subset of inflammatory bowel disease patients²², might in part function by preventing TNF-mediated necrosis of epithelial cells. Taken together, our findings revealed a previously unrecognized essential physiological function of FADD in protecting epithelial cells from RIP3-dependent necrosis and preventing intestinal inflammation *in vivo*. This function of FADD seems to be cell specific, as conditional FADD ablation did not sensitize hepatocytes²⁶ or oligodendrocytes²⁷ to spontaneous programmed necrosis but on the contrary protected these cells from TNF- or autoimmune-inflammation-induced cytotoxicity, respectively. In addition to recent studies showing that regulation of RIP-kinase-mediated necrosis is important for embryonic development^{28–30}, our findings provide a paradigm demonstrating that sensitization of epithelial cells to programmed necrosis triggers chronic inflammation *in vivo*, highlighting the significance of the mechanisms regulating programmed necrosis for the maintenance of physiological immune homeostasis and the prevention of inflammation in epithelial surfaces.

METHODS SUMMARY

Mice were maintained at the SPF animal facility of the Institute for Genetics, University of Cologne. Mice were either generated using gene targeting in C57BL/6 embryonic stem cells (Bruce4) or backcrossed for at least 10 generations

into the C57BL/6 genetic background. Germ-free mice were produced at the gnotobiotic facility of the University of Ulm and were conventionalized as described in detail in Supplementary Table 1. Endoscopic analysis was performed using a high-resolution mini-endoscope, Coloview (Karl-Storz). IECs were isolated by sequential incubation of intestinal tissue in 1 mM dithiothreitol (DTT) and 1.5 mM EDTA solutions. RNA preparation and RT-PCR analysis, protein extraction and immunoblotting, tissue preparation and immunohistological analysis were performed using standard protocols. Haematoxylin & eosin stained sections were scored in a blinded fashion for the amount of inflammation and tissue damage on separate scales from 0 to 3, which were added to a total score of 0 to 6.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 December 2010; accepted 6 June 2011.

Published online 31 July 2011.

- Kaser, A., Zeissig, S. & Blumberg, R. S. Inflammatory bowel disease. *Annu. Rev. Immunol.* **28**, 573–621 (2010).
- Strober, W., Fuss, I. & Mannon, P. The fundamental basis of inflammatory bowel disease. *J. Clin. Invest.* **117**, 514–521 (2007).
- MacDonald, T. T. & Monteleone, G. Immunity, inflammation, and allergy in the gut. *Science* **307**, 1920–1925 (2005).
- Xavier, R. J. & Podolsky, D. K. Unravelling the pathogenesis of inflammatory bowel disease. *Nature* **448**, 427–434 (2007).
- Turner, J. R. Intestinal mucosal barrier function in health and disease. *Nature Rev. Immunol.* **9**, 799–809 (2009).
- Wilson, N. S., Dixit, V. & Ashkenazi, A. Death receptor signal transducers: nodes of coordination in immune signaling networks. *Nature Immunol.* **10**, 348–355 (2009).
- Cho, Y. S. *et al.* Phosphorylation-driven assembly of the RIP1–RIP3 complex regulates programmed necrosis and virus-induced inflammation. *Cell* **137**, 1112–1123 (2009).
- He, S. *et al.* Receptor interacting protein kinase-3 determines cellular necrotic response to TNF- α . *Cell* **137**, 1100–1111 (2009).
- Zhang, D. W. *et al.* RIP3, an energy metabolism regulator that switches TNF-induced cell death from apoptosis to necrosis. *Science* **325**, 332–336 (2009).
- Hitomi, J. *et al.* Identification of a molecular signaling network that regulates a cellular necrotic cell death pathway. *Cell* **135**, 1311–1323 (2008).
- Nenci, A. *et al.* Epithelial NEMO links innate immunity to chronic intestinal inflammation. *Nature* **446**, 557–561 (2007).
- Zhang, J., Cado, D., Chen, A., Kabra, N. H. & Winoto, A. Fas-mediated apoptosis and activation-induced T-cell proliferation are defective in mice lacking FADD/Mort1. *Nature* **392**, 296–300 (1998).
- Yeh, W. C. *et al.* FADD: essential for embryo development and signaling from some, but not all, inducers of apoptosis. *Science* **279**, 1954–1958 (1998).
- Vandenabeele, P., Galluzzi, L., Vanden Berghe, T. & Kroemer, G. Molecular mechanisms of necroptosis: an ordered cellular explosion. *Nature Rev. Mol. Cell Biol.* **11**, 700–714 (2010).
- Holler, N. *et al.* Fas triggers an alternative, caspase-8-independent cell death pathway using the kinase RIP as effector molecule. *Nature Immunol.* **1**, 489–495 (2000).
- Osborn, S. L. *et al.* Fas-associated death domain (FADD) is a negative regulator of T-cell receptor-mediated necroptosis. *Proc. Natl Acad. Sci. USA* **107**, 13034–13039 (2010).
- Upton, J. W., Kaiser, W. J. & Mocarski, E. S. Virus inhibition of RIP3-dependent necrosis. *Cell Host Microbe* **7**, 302–313 (2010).
- Newton, K., Sun, X. & Dixit, V. M. Kinase RIP3 is dispensable for normal NF- κ Bs, signaling by the B-cell and T-cell receptors, tumor necrosis factor receptor 1, and Toll-like receptors 2 and 4. *Mol. Cell Biol.* **24**, 1464–1469 (2004).
- Kovalenko, A. *et al.* The tumour suppressor CYLD negatively regulates NF- κ B signalling by deubiquitination. *Nature* **424**, 801–805 (2003).
- Wang, L., Du, F. & Wang, X. TNF- α induces two distinct caspase-8 activation pathways. *Cell* **133**, 693–703 (2008).
- Kollias, G. TNF pathophysiology in murine models of chronic inflammation and autoimmunity. *Semin. Arthritis Rheum.* **34**, 3–6 (2005).
- Peyrin-Biroulet, L. Anti-TNF therapy in inflammatory bowel diseases: a huge review. *Minerva Gastroenterol. Dietol.* **56**, 233–243 (2010).
- Kaser, A. *et al.* XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell* **134**, 743–756 (2008).
- Cadwell, K. *et al.* A key role for autophagy and the autophagy gene *Atg16l1* in mouse and human intestinal Paneth cells. *Nature* **456**, 259–263 (2008).
- Dourmashin, R. R. *et al.* Epithelial patchy necrosis in Crohn's disease. *Hum. Pathol.* **14**, 643–648 (1983).
- Luedde, T. *et al.* Deletion of NEMO/IKK γ in liver parenchymal cells causes steatohepatitis and hepatocellular carcinoma. *Cancer Cell* **11**, 119–132 (2007).
- Mc Guire, C. *et al.* Oligodendrocyte-specific FADD deletion protects mice from autoimmune-mediated demyelination. *J. Immunol.* **185**, 7646–7653 (2010).
- Zhang, H. *et al.* Functional complementation between FADD and RIP1 in embryos and lymphocytes. *Nature* **471**, 373–376 (2011).
- Oberst, A. *et al.* Catalytic activity of the caspase-8-FLIP_L complex inhibits RIPK3-dependent necrosis. *Nature* **471**, 363–367 (2011).
- Kaiser, W. J. *et al.* RIP3 mediates the embryonic lethality of caspase-8-deficient mice. *Nature* **471**, 368–372 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Uthoff-Hachenberg, J. Pfeiffer, E. Mahlberg, D. Beier, J. Buchholz, B. Huelser, B. Wolff, E. Merkel and S. Schmidt for technical support. We also thank V. Dixit and K. Newton for providing *Ripk3*^{-/-} mice and R. Massoumi for providing anti-CYLD antibody. This work was funded by grants from the Deutsche Forschungsgemeinschaft (SFB 670, SFB 829, CECAD) and European Commission FP7 program grants 'INFLA-CARE' and 'Masterswitch' (EC contract numbers 223151 and 223404 respectively) to M.P. V. F.-M. was supported by a long-term EMBO fellowship, G.v.L. was supported by 'Group-ID MRP' of Ghent University and P.-S.W. was supported by a fellowship from the International Graduate School in Genetics and Functional Genomics at the University of Cologne.

Author Contributions P.-S.W., A.W., K.V., V.K., V.F.-M., M.E., P.K. and G.v.L. performed the research. P.-S.W., A.W., K.V., V.K., V.F.-M. and A.S.-K. analysed the data and contributed to writing the manuscript. M.P. provided ideas, co-ordinated the project and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.P. (Pasparakis@uni-koeln.de).

METHODS

Mice. FADD^{FL} mice were generated as described previously²⁷ and the CYLDA932^{FL} mice were generated as described in Supplementary Fig. 4, using gene targeting in C57BL/6 embryonic stem cells (Bruce 4). Villin-Cre (ref. 31); *Tnfr*^{-/-} (ref. 32), *Myd88*^{-/-} (ref. 33) and *Ripk3*^{-/-} (ref. 18) mice were backcrossed for at least 10 generations into the C57BL/6 genetic background. Mice were maintained at the SPF animal facility of the Institute for Genetics, University of Cologne, kept under a 12 h light cycle, and given a regular chow diet (Harlan, diet no. 2918) *ad libitum*. All animal procedures were conducted in accordance with European, national and institutional guidelines and protocols and were approved by local government authorities. For antibiotic treatment 1 g l⁻¹ ampicillin (ICN Biomedicals), 1 g l⁻¹ neomycin (Sigma), 0.5 g l⁻¹ meronem (AstraZeneca) and 0.5 g l⁻¹ ciprofloxacin (Fluka) were added to the drinking water starting from the second day after birth. At weaning, ciprofloxacin was substituted by 0.5 g l⁻¹ vancomycin (Eberth). Germ-free mice were produced at the gnotobiotic facility of the University of Ulm. Germ-free mice were conventionalized as described in detail in Supplementary Table 1. Sex-matched littermates not carrying the villin-Cre transgene were used as controls in all experiments. Unless otherwise indicated, mice were analysed between 6–12 weeks of age.

High-resolution mini-endoscopy. Mice were anaesthetized using intraperitoneal injection of ketamine (Ratiopharm)/rompun (Bayer) and a high-resolution mini-endoscope, Coloview (Karl-Storz), was used to determine the murine endoscopic index of colitis severity (MEICS), as described previously³⁴.

IEC isolation and immunoblotting. IECs were isolated by sequential incubation of intestinal tissue in 1 mM dithiothreitol (DTT) and 1.5 mM EDTA solutions as described previously³⁵. Protein extracts were prepared from IECs as described³⁶. Protease and phosphatase inhibitor tablets (Roche) were added to the lysis buffer. Protein extracts were separated by 10% SDS-PAGE gels and transferred to Immobilon-P PVDF membranes (Millipore). Membranes were probed with primary antibodies anti-FADD, anti- α -tubulin (Sigma), anti- β -actin (Santa Cruz), anti-mouse RIP3 (Enzo), anti-CYLD (provided by R. Masoumi). Membranes were incubated with secondary HRP-coupled antibodies (GE Healthcare and Jackson ImmuneResearch) and developed with chemiluminescent detection substrate (GE Healthcare and Thermo Scientific).

Histology. Tissues were fixed overnight in 4% paraformaldehyde, embedded in paraffin and cut in 4- μ m sections. Paraffin sections were rehydrated and heat-induced antigen retrieval was performed either in 10 mM sodium citrate, 0.05% Tween-20 pH 6 or in TEX (50 mM Tris, 1 mM EDTA, 0.5% Triton X-100; pH 8) with 20 μ g ml⁻¹ protease K. Primary antibodies used for IHC were anti-Ki67 (Dako), anti-Gr-1 (Pharmingen), anti-F4/80, anti-B220 (homemade), anti-CD3 (Abcam), anti-active caspase 3 (R&D systems), anti-human lysozyme (Dako). Biotinylated secondary antibodies were purchased from Perkin Elmer and Dako. Stainings were visualized with ABC Kit Vectastain Elite (Vector) and DAB substrate (DAKO). Incubation times with the DAB substrate were equal for all samples. Haematoxylin & eosin stained sections were scored in a blinded fashion for the amount of inflammation and tissue damage on separate scales from 0 to 3, which were added to obtain a total histological colitis score of 0 to 6. For inflammation the scoring was defined as follows: 0, no inflammatory infiltrate in the lamina propria; 1, increased presence of inflammatory cells in the mucosa; 2, inflammatory infiltrate extending into the submucosa; 3, transmural extension of inflammatory infiltrate. For tissue damage, the scoring was defined as follows: 0, no mucosal damage; 1, discrete epithelial lesions; 2, extended epithelial damage associated with areas containing elongated crypts, crypt abscesses or focal

ulceration; 3, extensive ulceration of the bowel wall. For scoring the amount of inflammation and tissue damage in the small intestine similar scales from 0 to 3 were used, which were added to obtain a total histological score of 0 to 6. For small intestinal inflammation the scoring was defined as follows: 0, no inflammatory infiltrate in the lamina propria; 1, increased presence of inflammatory cells between the crypts; 2, inflammatory infiltrate extending into the villi; 3, extension of inflammatory infiltrate throughout the lamina propria. For tissue damage, the scoring was based on the percentage of small intestinal crypts affected by IEC death and Paneth cell loss as follows: 0, 0–10% of crypts affected; 1, 10–40% of crypts affected; 2, 40–70% of crypts affected; 3, more than 70% of crypts affected. For electron microscopy 3-mm-long samples from distal, medial and proximal colon were excised from FADD^{FL} and FADD^{IEC-KO} mice and immediately fixed in 2.5% glutaraldehyde/2% paraformaldehyde in phosphate buffer pH 7.4 for 3 h at room temperature (20 °C). The tissues were post-fixed in 1% OsO₄ and embedded in Epon resin. 70–90-nm sections were cut, stained with uranyl acetate and lead citrate and viewed under a Philips CM-10 transmission electron microscope. Representative pictures were captured using an Orius SC200 CCD camera (Gatan GmbH). Evaluation of cell death on histological sections was performed by an experienced pathologist (A.S.-K.).

Cell death assays. Primary MEFs were isolated from wild-type and homozygous CYLDA932 mice. For the induction of necroptosis, cells were pre-treated for one hour with 1 μ g ml⁻¹ cycloheximide (CHX; Sigma) and 20 μ M zVAD-fmk (ENZO) and subsequently stimulated with 1, 10 or 30 ng ml⁻¹ murine TNF for 12 h. Cell survival was determined by spectrophotometric measurement of crystal violet incorporation. Values are presented as per cent survival of triplicates compared to untreated cells.

Quantitative RT-PCR. Total RNA was extracted with Trizol Reagent (Invitrogen) and RNeasy Columns (Qiagen) and cDNA was prepared with Superscript III cDNA-synthesis Kit (Invitrogen). RT-PCR was performed with SyBrGreen or TaqMan analysis (Applied Biosystems). TATA-box-binding protein was used as a reference gene.

Southern blotting. Genomic DNA extraction, digestion and Southern blotting were performed according to standard protocols. The probe used for Southern blot analysis of the *Fadd* locus was amplified using primers: sense 5'-CGTGAGGA GCAGGCAAGCAG-3' and antisense 5'-TGGTGAAGCCCTCCAGCCTGT-3'.

Statistics. All data shown represent the mean \pm s.d. Statistical analyses were performed with unpaired Student's *t*-tests with unequal variance. **P* \leq 0.05; ***P* \leq 0.01; ****P* \leq 0.005.

- Madison, B. B. *et al.* *cis* elements of the villin gene control expression in restricted domains of the vertical (crypt) and horizontal (duodenum, cecum) axes of the intestine. *J. Biol. Chem.* **277**, 33275–33283 (2002).
- Pasparakis, M., Alexopoulou, L., Episkopou, V. & Kollias, G. Immune and inflammatory responses in TNF α -deficient mice: a critical requirement for TNF α in the formation of primary B cell follicles, follicular dendritic cell networks and germinal centers, and in the maturation of the humoral immune response. *J. Exp. Med.* **184**, 1397–1411 (1996).
- Adachi, O. *et al.* Targeted disruption of the *MyD88* gene results in loss of IL-1- and IL-18-mediated function. *Immunity* **9**, 143–150 (1998).
- Becker, C. *et al.* *In vivo* imaging of colitis and colon cancer development in mice using high resolution chromoendoscopy. *Gut* **54**, 950–954 (2005).
- Ukena, S. N. *et al.* Probiotic *Escherichia coli* Nissle 1917 inhibits leaky gut by enhancing mucosal integrity. *PLoS ONE* **2**, e1308 (2007).
- Schmidt-Suppran, M. *et al.* NEMO/IKK γ -deficient mice model incontinentia pigmenti. *Mol. Cell* **5**, 981–992 (2000).

Caspase-8 regulates TNF- α -induced epithelial necroptosis and terminal ileitis

Claudia Günther¹, Eva Martini¹, Nadine Wittkopf¹, Kerstin Amann², Benno Weigmann¹, Helmut Neumann¹, Maximilian J. Waldner¹, Stephen M. Hedrick³, Stefan Tenzer⁴, Markus F. Neurath¹ & Christoph Becker¹

Dysfunction of the intestinal epithelium is believed to result in the excessive translocation of commensal bacteria into the bowel wall that drives chronic mucosal inflammation in Crohn's disease, an incurable inflammatory bowel disease in humans characterized by inflammation of the terminal ileum¹. In healthy individuals, the intestinal epithelium maintains a physical barrier, established by the tight contact of cells. Moreover, specialized epithelial cells such as Paneth cells and goblet cells provide innate immune defence functions by secreting mucus and antimicrobial peptides, which hamper access and survival of bacteria adjacent to the epithelium². Epithelial cell death is a hallmark of intestinal inflammation and has been discussed as a possible pathogenic mechanism driving Crohn's disease in humans³. However, the regulation of epithelial cell death and its role in intestinal homeostasis remain poorly understood. Here we demonstrate a critical role for caspase-8 in regulating necroptosis of intestinal epithelial cells (IECs) and terminal ileitis. Mice with a conditional deletion of caspase-8 in the intestinal epithelium (*Casp8*^{ΔIEC}) spontaneously developed inflammatory lesions in the terminal ileum and were highly susceptible to colitis. *Casp8*^{ΔIEC} mice lacked Paneth cells and showed reduced numbers of goblet cells, indicating dysregulated antimicrobial immune cell functions of the intestinal epithelium. *Casp8*^{ΔIEC} mice showed increased cell death in the Paneth cell area of small intestinal crypts. Epithelial cell death was induced by tumour necrosis factor (TNF)- α , was associated with increased expression of receptor-interacting protein 3 (*Rip3*; also known as *Ripk3*) and could be inhibited on blockade of necroptosis. Lastly, we identified high levels of RIP3 in human Paneth cells and increased necroptosis in the terminal ileum of patients with Crohn's disease, suggesting a potential role of necroptosis in the pathogenesis of this disease. Together, our data demonstrate a critical function of caspase-8 in regulating intestinal homeostasis and in protecting IECs from TNF- α -induced necroptotic cell death.

Caspase-8 is a cysteine protease critically involved in regulating cellular apoptosis. On activation of death receptors, including TNF-receptor and Fas, caspase-8 is activated by limited autoproteolysis and the processed caspase-8 subsequently triggers the caspase cascade that finally leads to apoptotic cell death. Caspase-mediated apoptosis is important for the turnover of IECs and for shaping the morphology of the gastrointestinal tract⁴. Furthermore, recent data have indicated a role of caspase-mediated apoptosis of IECs in the pathogenesis of inflammatory bowel diseases (IBDs) such as Crohn's disease and ulcerative colitis^{5,6}.

To study the function of caspase-8 in the gut, we generated mice with an IEC-specific deletion of caspase-8 (*Casp8*^{ΔIEC}). Accordingly, mice with floxed caspase-8 alleles were bred with mice expressing the Cre recombinase under the control of the IEC-specific villin promoter. Specific deletion of caspase-8 in IECs was confirmed by polymerase chain reaction (PCR) and western blotting (Supplementary Fig. 1a, b). *Casp8*^{ΔIEC} mice were born at the expected Mendelian ratios and

developed normally, although weighing on average slightly less than control littermates at 8 weeks of age (data not shown). Despite the paradigm that apoptosis is important for regulating epithelial cell numbers⁴, histological and morphometrical analysis of the jejunum and colon of *Casp8*^{ΔIEC} mice showed no overt changes of tissue architecture or dysregulation of apoptosis (Supplementary Figs 1d–f and 2). Although this suggested that caspase-8 is not essential for the structural development of the gut, high-resolution endoscopy showed erosions in the terminal ileum—but not in the colon—of *Casp8*^{ΔIEC} mice (Fig. 1a and Supplementary Fig. 1c). Histological analysis demonstrated marked destruction of the architecture and signs of inflammation including bowel wall thickening, crypt loss and increased cellularity in the lamina propria (Fig. 1b) in more than 80% of all ileal specimens. This finding of spontaneous ileitis in the absence of caspase-8 in IECs was further supported by increased expression of the inflammation markers *S100a9* and *TNF- α* (also known as *Tnf*) and by elevated infiltration of the lamina propria with CD4⁺ T cells and granulocytes (Fig. 1c and Supplementary Fig. 3).

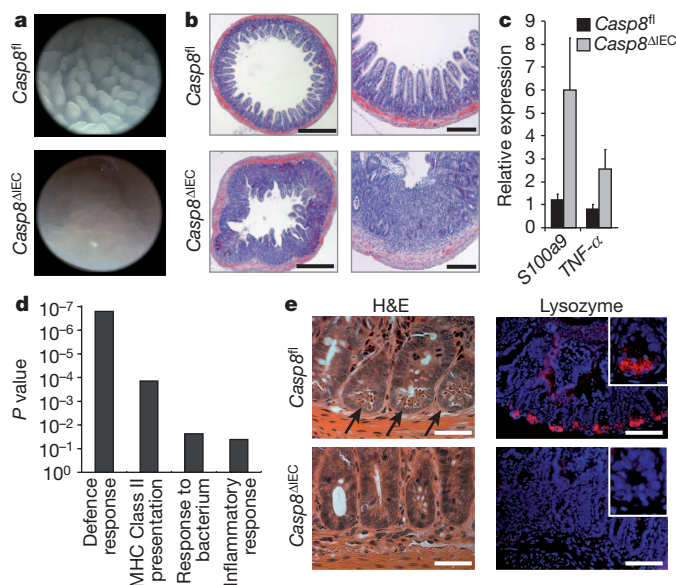


Figure 1 | *Casp8*^{ΔIEC} mice spontaneously develop ileitis and lack Paneth cells. **a**, **b**, Representative endoscopic pictures (**a**) and H&E stained cross-sections (**b**) showing villous erosions in the terminal ileum of *Casp8*^{ΔIEC}. Scale bars, 500 μ m (left), 200 μ m (right). **c**, RT-PCR showing increased level of inflammatory markers in the terminal ileum of *Casp8*^{ΔIEC} mice (6 mice per group + s.e.m., relative to *Hprt*). **d**, GO analysis of genes significantly downregulated in gene-chip analysis of IECs from three control and three *Casp8*^{ΔIEC} mice. **e**, Ileum cross-sections stained with H&E (scale bars, 50 μ m) and lysozyme (scale bars, 100 μ m) for Paneth cells (inset shows single crypt at higher magnification). Arrows indicate crypt bottom with Paneth cells.

¹Department of Medicine 1, Friedrich-Alexander-University, D-91054 Erlangen, Germany. ²Department of Nephropathology, Friedrich-Alexander-University, D-91054 Erlangen, Germany. ³Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093, USA. ⁴Institute of Immunology, Johannes Gutenberg University, D-55131 Mainz, Germany.

To investigate whether caspase-8 deficiency sensitizes mice to experimental intestinal inflammation in the large intestine, we subjected *Casp8^{ΔIEC}* and control mice to dextran sodium sulphate, a well-established model of experimental colitis. Notably, we observed high lethality in the group of *Casp8^{ΔIEC}* mice but not in control mice (Supplementary Fig. 4). Moreover, the former lost significantly more weight than controls. All *Casp8^{ΔIEC}* but none of the control mice developed rectal bleeding and endoscopic and histological signs of very severe colitis with epithelial erosions, a finding that was confirmed by quantitative PCR for the IEC marker villin. Together, our data indicate that a lack of caspase-8 in IECs renders mice highly susceptible to spontaneous ileitis and experimentally induced colitis.

To screen for molecular mechanisms that sensitize *Casp8^{ΔIEC}* mice for intestinal inflammation, we next performed whole-genome gene-chip analysis of IECs from unchallenged control and *Casp8^{ΔIEC}* mice. Of the 45,000 expression tags analysed, 197 were significantly downregulated and 136 upregulated in *Casp8^{ΔIEC}* mice (fold change >2.0, $P < 0.05$) (Supplementary Table 1). Gene ontology (GO) analysis of downregulated genes showed that several biological pathways were impaired in *Casp8^{ΔIEC}* mice as compared to littermate controls, with the GO terms “defence response” and “MHC class II antigen presentation” reaching very high levels of significance (Fig. 1d). Within the group of genes upregulated in *Casp8^{ΔIEC}* mice, no GO term reached significance levels. Notably, within the defence response gene set, genes belonging to the family of antimicrobial peptides including several defensins, lysozyme and phospholipases were among the most strongly downregulated genes (Supplementary Fig. 5a), suggesting defects in the antimicrobial defence of the intestinal epithelium in the absence of epithelial caspase-8. Expression of antimicrobial peptides is a hallmark of Paneth cells, epithelial-derived cells that are located at the base of the crypts of Lieberkuhn in the small intestine⁷. Within Paneth cells, antimicrobial peptides are stored in cytoplasmic granules, from which they can be released into the gut lumen, thereby contributing to intestinal host defences. Notably, *Casp8^{ΔIEC}* mice showed a complete absence of cells with secretory granules and lysozyme expression at the crypt base of the small intestine, suggesting that Paneth cells are lacking in the gut of these animals (Fig. 1e). Furthermore, the number of mucus-secreting goblet cells was reduced, as indicated by staining with ulex europaeus agglutinin 1 (UEA-1), a lectin binding to glycoproteins characteristic for these cells (Supplementary Fig. 5b). In contrast, we observed no changes in the appearance of enteroendocrine cells or absorptive enterocytes, as indicated by staining for chromogranin-A or alkaline phosphatase, respectively (Supplementary Fig. 5b). The lack of Paneth cells and partial lack of goblet cells was confirmed by quantitative gene expression analysis showing diminished expression of Paneth-cell- and goblet-cell-specific genes, whereas expression of genes specific for enteroendocrine cells, enterocytes and progenitor cells was unchanged (Supplementary Fig. 5c). Thus, collectively, our data indicate that deficient expression of caspase-8 in the intestinal epithelium results in diminished Paneth and goblet cell numbers and hence may lead to defects in antimicrobial host defence and terminal ileitis.

In addition to controlling apoptosis, there is growing evidence that caspase-8 regulates several non-apoptotic cellular mechanisms including proliferation, migration and differentiation^{8,9}. Accordingly, caspase-8 has been shown to promote the terminal differentiation of macrophages and keratinocytes¹⁰. Thus, we reasoned that caspase-8 might support intestinal immune homeostasis by promoting the terminal differentiation of Paneth and goblet cells. To verify this hypothesis, we performed long-term organoid cultures of small intestinal crypts *in vitro* as previously described¹¹. Isolated crypts from the small intestine underwent multiple crypt fissions forming large organoids in both control and *Casp8^{ΔIEC}* mice (Supplementary Fig. 6a). In marked contrast to the absence of Paneth cells in crypts from *Casp8^{ΔIEC}* mice *in vivo*, organoids grown from these crypts *in vitro* over a period of 1–4 weeks showed Paneth cells indistinguishable in localization and number from organoids cultured from control littermate mice (Fig. 2a).

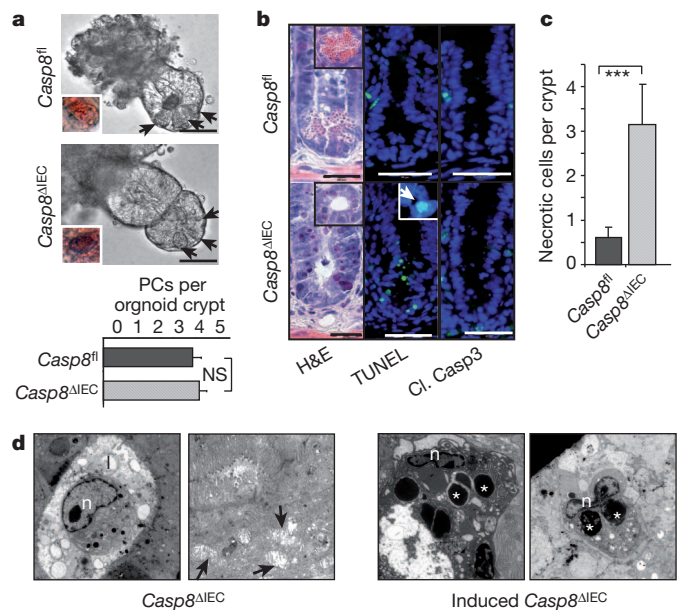


Figure 2 | Increased caspase-8-independent cell death within crypts of *Casp8^{ΔIEC}* mice. **a**, Representative pictures of gut organoids (scale bars, 50 μ m). Arrows indicate Paneth cells. Insets show eosin staining indicating Paneth cells. Graph shows number of Paneth cells (PCs) per organoid crypt ($n = 24$) + s.e.m. NS, not significant. **b**, Crypt cross-sections from the small intestine of control and *Casp8^{ΔIEC}* mice stained with H&E (scale bars, 20 μ m), TUNEL (inset shows condensed nuclei) and cleaved caspase-3 (Cl. Casp3) (scale bars, 50 μ m). **c**, Quantification of necrotic cells per crypt + s.d. of control ($n = 9$) and *Casp8^{ΔIEC}* ($n = 14$) mice. **d**, Electron microscopic pictures of dying crypt cells in *Casp8^{ΔIEC}* mice and inducible *Casp8^{ΔIEC}*. Asterisks indicate Paneth cell granules; arrows indicate mitochondrial swelling. l, crypt lumen; n, nuclei.

PCR analysis confirmed deletion of the caspase-8 allele in Paneth-cell-positive organoids derived from *Casp8^{ΔIEC}* mice (Supplementary Fig. 6b). Thus, our data indicate that caspase-8 is not required for the differentiation of IECs into Paneth cells, but that a factor present *in vivo* either inhibited the development of or ablated Paneth cells in the absence of caspase-8 expression.

Indeed, *Casp8^{ΔIEC}* mice showed a large number of dying epithelial cells at the crypt base with pyknotic nuclei and a shrunken eosinophilic cytoplasm (Fig. 2b), implying that caspase-8-deficient Paneth and goblet cells might be sensitive to cell death. Dying crypt cells usually lacked typical apoptotic body formation, suggesting necrotic rather than apoptotic cell death. This conclusion was supported by the observation that dying cells were TdT-mediated dUTP nick end labelling (TUNEL) positive, but showed no activation of caspase-3 (Fig. 2b). The number of necrotic cells at the crypt base was significantly higher in *Casp8^{ΔIEC}* mice as compared to control mice (Fig. 2c). Lastly, electron microscopy of the crypt area demonstrated cells with typical features of necrosis including mitochondrial swelling and extensive vacuole formation while typically lacking the blebbing usually associated with apoptosis (Fig. 2d). Importantly, many cells with features of necrosis also showed electron-dense granules, indicating necrotic Paneth cells. This conclusion was supported by electron microscopy of mice in which the caspase-8 deletion was induced in adult mice by injection with tamoxifen (inducible *Casp8^{ΔIEC}*) to detect early effects of caspase-8 deletion. Taken together, our data indicate that the lack of caspase-8 sensitizes Paneth cells in the crypts of the small intestine to necrotic cell death.

TNF- α -stimulated death receptor signalling has been described to promote necrosis in a number of different target cell types, especially when apoptosis was blocked using caspase inhibitors^{12,13}. We therefore reasoned that in the absence of caspase-8, TNF- α signalling might lead to excessive crypt cell death. To test this hypothesis, we intravenously administered TNF- α to *Casp8^{ΔIEC}* mice using a dose that is not lethal

to normal mice. Whereas all control mice were still alive after 5 h, *Casp8^{ΔIEC}* mice showed significantly more pronounced hypothermia and very high lethality (Supplementary Fig. 7a, b). Histological analysis demonstrated villous atrophy and severe destruction of the small bowel of *Casp8^{ΔIEC}* mice as compared to control littermates and an increased number of dying epithelial cells, as indicated by the pyknotic nuclei seen in the haematoxylin and eosin (H&E) stain of crypts and cells in the crypt lumen (Supplementary Fig. 7c–e). Similar to unchallenged mice, dying crypt cells were negative for active caspase-3 but positive for TUNEL staining (Supplementary Fig. 8a–c), suggesting that in the absence of caspase-8, TNF drives excessive necrosis of epithelial cells.

Recent data from other experimental systems have shown that inhibition of caspase activity in genetic models or by using specific caspase inhibitors can result in an apoptosis-independent type of programmed necrosis called necroptosis¹⁴. Necroptosis has been shown to be mediated by the kinases RIP1 and RIP3 (refs 15, 16). On induction of necroptosis, RIP3 is recruited to RIP1 to establish a necroptosis-inducing protein complex. RIP3 seems to be essential for the molecular mechanisms driving necroptosis and expression of *Rip3* has been demonstrated to correlate with the sensitivity of cells towards necroptosis^{17,18}. Moreover, deletion of *Rip3* has recently been shown to rescue the lethal phenotype of general caspase-8-deficient mice by blocking cell death^{19,20}. Notably, expression of *Rip3* messenger RNA, but not *Rip1* (also known as *Ripk1*) mRNA was significantly increased in IECs isolated from unchallenged *Casp8^{ΔIEC}* mice as compared to controls (Supplementary Fig. 9). Condensed nuclei as observed in the crypts of *Casp8^{ΔIEC}* mice stained for RIP3 using immunohistochemistry (Fig. 3a and Supplementary Fig. 8e). Moreover, RIP3 was overexpressed in the small intestine of TNF- α -treated *Casp8^{ΔIEC}* mice when compared to control littermate mice (Fig. 3b and Supplementary Fig. 8d), suggesting that the lack of caspase-8 in the intestinal epithelium might sensitize IECs to RIP-mediated necroptotic cell death. In line with this hypothesis, RIP3 staining was detected especially in cells at the crypt

base (Supplementary Fig. 10). Moreover, significant levels of TNF- α were detected in lamina propria cells adjacent to crypt IECs and both TNF- α and *Rip3* expression were highest in the terminal ileum.

RIP-mediated necroptosis can be blocked *in vitro* and *in vivo* by using necrostatin-1 (nec-1), an allosteric small-molecule inhibitor of the RIP1 kinase²¹. Thus we reasoned that nec-1 might prevent TNF- α -induced epithelial necroptosis and lethality in *Casp8^{ΔIEC}* mice. Indeed, *in vitro*, small intestinal organoid cultures from *Casp8^{ΔIEC}* mice but not from control mice exhibited necrosis within 24 h after addition of TNF- α to the tissue culture. However, when cell cultures were pre-treated with nec-1, organoid necrosis was blocked (Fig. 3c, d). Moreover, pre-treatment with nec-1 significantly reduced TNF- α -induced lethality and small intestinal tissue destruction in *Casp8^{ΔIEC}* mice (Fig. 3e, f). Collectively, our data indicate that deficient caspase-8 expression renders Paneth cells susceptible to TNF- α -induced necroptosis, highlighting a regulatory role of caspase-8 in antimicrobial defence and in maintaining immune homeostasis in the gut.

Interestingly, defects in Paneth cell function and in the expression of antimicrobial peptides have been described in Crohn's disease patients and accumulating evidence supports a role for Paneth cells in the pathogenesis of this disease^{1,7,22}. As anti-TNF- α treatment is successfully used in the therapy of patients with Crohn's disease, we hypothesized that, similar to *Casp8^{ΔIEC}* mice, human Paneth cells might be susceptible to TNF- α -induced necroptosis. Paneth cell dysfunction has been reported in mice deficient of autophagy genes such as *Atg16l1*, a gene associated with Crohn's disease susceptibility³. Moreover, depletion of Paneth cells in the small intestine has been reported in patients with ileal Crohn's disease and patients with ulcerative colitis involving the terminal ileum²³. Notably, immunohistochemistry of samples derived from the terminal ileum of human patients undergoing endoscopic examination revealed high expression of RIP3 exclusively in Paneth cells (Fig. 4a), but not in other intestinal epithelial cell types. Importantly, analysis of histological samples from the terminal ileum of control patients and patients with active Crohn's disease showed a significant decrease in the number of Paneth cells and high numbers of dying cells with shrunken eosinophilic cytoplasm (Fig. 4b) at the crypt base, similar to *Casp8^{ΔIEC}* mice. Electron microscopy of the Paneth cell area in the terminal ileum of patients with Crohn's disease showed increased necrotic cell death, as indicated by abundant organelle swelling, vacuole formation and the lack of blebbing (Fig. 4c, d). Moreover, crypt epithelial cells in areas of acute inflammation usually were TUNEL positive but lacked staining for active caspase-3 (Fig. 4e). Lastly, ileal biopsies from control patients showed Paneth cell loss in the presence of high levels of exogenous TNF- α , an effect that was reversible by co-incubation with nec-1 (Fig. 4f and Supplementary Fig. 11a). Thus, our data indicate that necroptosis of Paneth cells is a feature of Crohn's disease. As it has recently been shown that anti-TNF treatment partially restores the deficient expression of antimicrobial peptides in Crohn's disease patients²⁴, our data indicate that the high levels of TNF- α present in the lamina propria of the inflamed ileum induce Paneth cell necroptosis and may provide a molecular explanation for the defects in antimicrobial defence observed in these patients.

Our data uncover an unexpected function of caspase-8 in regulating necroptosis of intestinal epithelial cells and in maintaining immune homeostasis in the gut. Caspase-8-deficient mice had no defect in overall gut morphology, demonstrating that cell death independent from the extrinsic apoptosis pathway can regulate intestinal homeostasis. Indeed, studies using electron microscopy have shown various different cell death morphologies in the small intestine including morphological changes usually seen in necrosis, such as cell swelling and a degradation of organelles and membranes²⁵. Caspase-8-deficient mice completely lacked Paneth cells, suggesting that these cell types are highly susceptible to necroptosis. Crohn's disease patients frequently show reduced Paneth and goblet cell numbers and reduced expression of Paneth-cell-derived defensins in areas of acute inflammation, suggesting that necroptosis

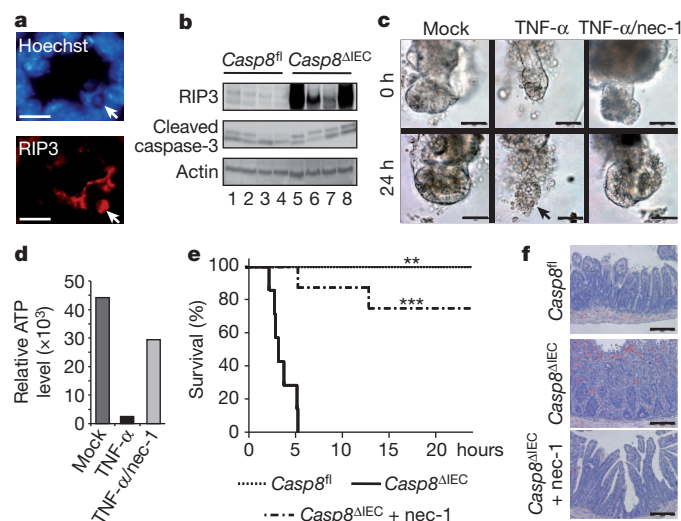


Figure 3 | Inhibition of TNF- α -induced epithelial necroptosis in *Casp8^{ΔIEC}* mice. **a**, Representative RIP3 staining co-localizing with condensed nuclei (arrows) at the crypt bottom of *Casp8^{ΔIEC}* mice. Scale bars, 20 μ m. **b**, Western blot for RIP3 and cleaved caspase-3 of IEC lysates isolated from TNF- α -treated control and *Casp8^{ΔIEC}* mice. Actin served as a control. **c**, **d**, Representative microscopic pictures (**c**) and cell viability (**d**) of *Casp8^{ΔIEC}* organoids treated for 24 h with TNF- α with or without nec-1. Scale bars, 50 μ m. Arrow indicates necrotic organoid. **e**, **f**, Survival (**e**) and H&E stained small intestine cross-sections of control ($n = 5$), *Casp8^{ΔIEC}* (mock pre-treated, $n = 7$) and *Casp8^{ΔIEC}* (nec-1 pretreated, $n = 8$) mice (**f**) after intravenous injection of TNF- α . All experiments were performed at least 3 times with similar results. Scale bars, 200 μ m. ** $P < 0.01$, *** $P < 0.001$, relative to *Casp8^{ΔIEC}* without nec-1.

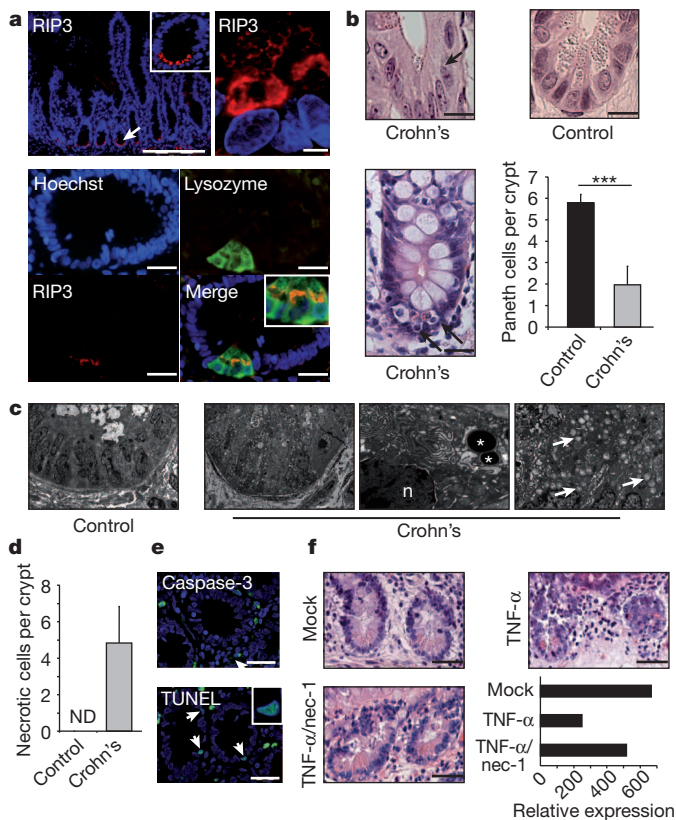


Figure 4 | RIP-mediated necroptosis of Paneth cells in patients with Crohn's disease. **a**, Representative RIP3 immunostaining of the terminal ileum (healthy patient). Top, RIP3 expression in human Paneth cells (scale bars, 200 μ m (left), 5 μ m (right)). Bottom, co-localization of lysozyme and RIP3 in Paneth cells (scale bars, 20 μ m). Arrow indicates Paneth cells. **b**, H&E staining of crypts in the terminal ileum. Arrows indicate crypt cells with shrunken eosinophilic cytoplasm and pyknotic nuclei (scale bars, 20 μ m). Graph shows number of Paneth cells (\pm s.d.) per crypt in control patients ($n = 7$) and patients with active Crohn's disease ($n = 4$). $***P < 0.001$. **c**, Electron microscopy of the terminal ileum of control and Crohn's disease patient. Asterisks highlight Paneth cell granules; arrows indicate mitochondrial swelling. n, nucleus. **d**, Number of crypt cells (\pm s.d.) showing organelle swelling but regular nuclei as signs of necroptosis. Electron microscope pictures of four patients were analysed. ND, not detectable. **e**, Representative immunofluorescence staining for TUNEL and cleaved caspase-3 in crypts of the terminal ileum of a Crohn's disease patient (scale bars, 50 μ m). Arrows indicate Paneth cells. **f**, H&E staining of biopsies from the small intestine of control patients stimulated *in vitro* with either DMSO (mock), TNF- α alone or in combination with nec-1 (scale bars, 50 μ m). Graph shows quantitative expression level of the Paneth cell marker lysozyme relative to *HPRT*. Data from one representative experiment out of two is shown.

might be involved in the pathogenesis of human IBD²². Indeed, we were able to demonstrate constitutive expression of RIP3—a kinase sensitizing cells to necroptosis^{17,18}—in human Paneth cells. Moreover, cells undergoing necroptosis were found at the crypt base in patients with Crohn's disease and Paneth cell death could be inhibited by blocking necroptosis.

Caspase-8 has recently been shown to suppress RIP3–RIP1-kinase-dependent necroptosis following death receptor activation^{14,17,18}. This has been highlighted by genetic studies demonstrating that deletion of *Rip3* can rescue the embryonic lethality observed in mice with a general deletion of caspase-8 (refs 19, 20). Thus it is becoming increasingly clear that caspase-8 has an essential function in controlling RIP3-mediated necroptosis. On the molecular level, caspase-8 has been demonstrated to proteolytically cleave and inactivate RIP1 and RIP3, thereby regulating the initiation of necroptosis²⁰. So far, to our knowledge, no study has demonstrated caspase-8 as an IBD-linked gene using genetic

studies. However, caspase-8 was expressed at relatively low levels in the crypt area of the human terminal ileum and activation of caspase-8 was only seen sporadically along the crypt–villous axis (Supplementary Fig. 11b). Given the low expression of caspase-8 and high expression of RIP3 at the base of crypts, cells residing in this area may be susceptible to necroptosis on stimulation with death receptor ligands such as TNF- α .

TNF- α is an important contributor to the pathogenesis of IBD and treatment with biological drugs targeting TNF- α is effective in patients with Crohn's disease. Interestingly, mice overproducing TNF- α develop transmural intestinal inflammation with granulomas primarily in the terminal ileum, similar to Crohn's disease²⁶. TNF- α was a strong promoter of intestinal epithelial necroptosis in our experiments. Because anti-TNF treatment has been shown to restore expression of certain antimicrobial peptides²⁴ it is tempting to speculate that TNF- α -induced necroptosis of intestinal epithelial cells contributes to the pathogenesis of IBD and that Paneth and goblet cell sensitivity towards TNF-induced necroptosis may be an early event in IBD development. However, it remains to be determined whether Paneth cell necroptosis in Crohn's disease patients is quantitatively sufficient to affect the disease process. Further studies will be needed to decipher the precise regulatory network of death receptors, RIP kinases and caspase-8 at the crypt base and it will be important to elucidate whether genetic, epigenetic or post-translational mechanisms restrict expression or activation of caspase-8 in IBD patients. Here, our data for the first time demonstrate necroptosis in the terminal ileum of patients with Crohn's disease and indicate that regulating necroptosis in the intestinal epithelium is critical for the maintenance of intestinal immune homeostasis. Targeting necroptotic cellular mechanisms emerges as a promising option in treating patients with IBD.

METHODS SUMMARY

Mice carrying a *loxP*-flanked caspase-8 allele (*Casp8^{fl}*) and villin-*Cre* mice were described earlier^{27–29}. IEC-specific caspase-8 knockout mice were generated by breeding *Casp8^{fl}* mice to villin-*Cre* or villin-*CreERT2* mice. Experimental colitis was induced with 1–1.5% dextran sodium sulphate (DSS; MP Biomedicals) in the drinking water for 5–10 days. Colitis development was monitored by analysis of weight, rectal bleeding and colonoscopy as previously described³⁰. In some experiments, mice were injected intravenously with TNF- α (200 ng g⁻¹ body weight; Immunotools) plus or minus nec-1 (1.65 μ g g⁻¹ body weight; Enzo). Histopathological analysis was performed on formalin-fixed paraffin-embedded tissue after H&E staining. Immunofluorescence of cryosections was performed using the TSA-Kit as recommended by the manufacturer (PerkinElmer). For electron microscopy, glutaraldehyde-fixed material was used. Ultrathin sections were cut and analysed using a Zeiss EM 906. Paraffin-embedded patient specimens were obtained from the Institute of Pathology and endoscopic biopsies were collected in the Department of Medicine 1 (Erlangen University). The collection of samples was approved by the local ethical committee and each patient gave written informed consent. For organoid culture, intestinal crypts were isolated from mice and cultured as previously described¹¹. Organoid growth was monitored by light microscopy. IECs were isolated as previously described⁵. For gene-chip experiments, total RNA of IECs was isolated using the RNeasy Mini Kit (Qiagen) and hybridized to the Affymetrix mouse 430 2.0 chip (Affymetrix). GO-based analyses were performed using the online tool Database for Annotation, Visualization and Integrated Discovery (DAVID). Caspase-3/-7 activity was measured using the Caspase-Glo3/7 Assay (Promega) according to the manufacturer's instructions.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 February; accepted 2 August 2011.

1. Strober, W., Fuss, I. & Mannon, P. The fundamental basis of inflammatory bowel disease. *J. Clin. Invest.* **117**, 514–521 (2007).
2. Artis, D. Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. *Nature Rev. Immunol.* **8**, 411–420 (2008).
3. Kaser, A., Zeissig, S. & Blumberg, R. S. Inflammatory bowel disease. *Annu. Rev. Immunol.* **28**, 573–621 (2010).
4. Hall, P. A. *et al.* Regulation of cell number in the mammalian gastrointestinal tract: the importance of apoptosis. *J. Cell Sci.* **107**, 3569–3577 (1994).
5. Nenci, A. *et al.* Epithelial NEMO links innate immunity to chronic intestinal inflammation. *Nature* **446**, 557–561 (2007).

6. Sanders, D. S. Mucosal integrity and barrier function in the pathogenesis of early lesions in Crohn's disease. *J. Clin. Pathol.* **58**, 568–572 (2005).
7. Elphick, D. A. & Mahida, Y. R. Paneth cells: their role in innate immunity and inflammatory disease. *Gut* **54**, 1802–1809 (2005).
8. Maelfait, J. & Beyaert, R. Non-apoptotic functions of caspase-8. *Biochem. Pharmacol.* **76**, 1365–1373 (2008).
9. Valmiki, M. G. & Ramos, J. W. Death effector domain-containing proteins. *Cell. Mol. Life Sci.* **66**, 814–830 (2009).
10. Mielgo, A. *et al.* The death effector domains of caspase-8 induce terminal differentiation. *PLoS ONE* **4**, e7879 (2009).
11. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
12. Holler, N. *et al.* Fas triggers an alternative, caspase-8-independent cell death pathway using the kinase RIP as effector molecule. *Nature Immunol.* **1**, 489–495 (2000).
13. Vercammen, D. *et al.* Tumour necrosis factor-induced necrosis versus anti-Fas-induced apoptosis in L929 cells. *Cytokine* **9**, 801–808 (1997).
14. Vandenabeele, P. *et al.* Molecular mechanisms of necroptosis: an ordered cellular explosion. *Nature Rev. Mol. Cell Biol.* **11**, 700–714 (2010).
15. Cho, Y. S. *et al.* Phosphorylation-driven assembly of the RIP1–RIP3 complex regulates programmed necrosis and virus-induced inflammation. *Cell* **137**, 1112–1123 (2009).
16. Declercq, W., Vanden Berghe, T. & Vandenabeele, P. RIP kinases at the crossroads of cell death and survival. *Cell* **138**, 229–232 (2009).
17. He, S. *et al.* Receptor interacting protein kinase-3 determines cellular necrotic response to TNF- α . *Cell* **137**, 1100–1111 (2009).
18. Zhang, D. W. *et al.* RIP3, an energy metabolism regulator that switches TNF-induced cell death from apoptosis to necrosis. *Science* **325**, 332–336 (2009).
19. Kaiser, W. J. *et al.* RIP3 mediates the embryonic lethality of caspase-8-deficient mice. *Nature* **471**, 368–372 (2011).
20. Oberst, A. *et al.* Catalytic activity of the caspase-8–FLIP_L complex inhibits RIPK3-dependent necrosis. *Nature* **471**, 363–367 (2011).
21. Degterev, A. *et al.* Identification of RIP1 kinase as a specific cellular target of necrostatins. *Nature Chem. Biol.* **4**, 313–321 (2008).
22. Wehkamp, J. *et al.* Barrier dysfunction due to distinct defensin deficiencies in small intestinal and colonic Crohn's disease. *Mucosal Immunol.* **1** (Suppl. 1), S67–S74 (2008).
23. Lewin, K. The Paneth cell in disease. *Gut* **10**, 804–811 (1969).
24. Ajijs, I. *et al.* Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment. *PLoS ONE* **4**, e7984 (2009).
25. Mayhew, T. M. *et al.* Epithelial integrity, cell death and cell loss in mammalian small intestine. *Histol. Histopathol.* **14**, 257–267 (1999).
26. Kontoyannis, D. *et al.* Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: implications for joint and gut-associated immunopathologies. *Immunity* **10**, 387–398 (1999).
27. Beisner, D. R. *et al.* Cutting edge: innate immunity conferred by B cells is regulated by caspase-8. *J. Immunol.* **175**, 3469–3473 (2005).
28. Madison, B. B. *et al.* *cis* elements of the villin gene control expression in restricted domains of the vertical (crypt) and horizontal (duodenum, cecum) axes of the intestine. *J. Biol. Chem.* **277**, 33275–33283 (2002).
29. El Marjou, F. *et al.* Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* **39**, 186–193 (2004).
30. Becker, C. *et al.* *In vivo* imaging of colitis and colon cancer development in mice using high resolution chromoendoscopy. *Gut* **54**, 950–954 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The research leading to these results has received funding from the Interdisciplinary Center for Clinical Research (IZKF) of the University Erlangen-Nuremberg and the European Community's 7th Framework Program (FP7/2007–2013) under grant agreement no. 202230, acronym GENINCA. E.M. received funding from the Wellcome Trust (WT087768MA) and S.M.H. was supported by NIH grant AI037988. The authors thank A. Watson for critical reading of the manuscript, A. Nikolaev, S. Wallmüller, V. Buchert and M. Klewer for technical assistance and J. Mudter, R. Atreya and C. Neufert for sampling biopsies.

Author Contributions C.G., K.A., M.F.N. and C.B. designed the research. C.G., E.M., N.W., B.W., H.N., M.W. and S.T. performed the experiments. S.M.H. provided material that made the study possible. C.G., K.A. and C.B. analysed the data and wrote the paper.

Author Information Chip data were deposited at the NCBI Gene Expression Omnibus under the series accession number GSE30873 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30873>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.B. (christoph.becker@uk-erlangen.de).

METHODS

Mice. Mice carrying a *loxP*-flanked caspase-8 allele (*Casp8^{fl}*) and villin-*Cre* mice were described earlier^{27–29}. Intestinal-epithelium-specific caspase-8 knockout mice were generated by breeding floxed caspase-8 mice to either villin-*Cre* or villin-*CreERT2* mice. For the induction of the *CreERT2* line (villin-*CreERT2* × *Casp8^{fl}*), tamoxifen (50 mg ml^{−1} ethanol; Sigma) was emulsified in sunflower oil at a concentration of 5 mg ml^{−1}. Mice were injected daily intraperitoneally with 200 µl of tamoxifen. In all experiments, littermates carrying the *loxP*-flanked alleles but not expressing *Cre* recombinase were used as controls. *Cre*-mediated recombination was genotyped by PCR on tail DNA. Experimental colitis was induced by treating mice with 1–1.5% dextrane sodium sulphate (DSS; MP Biomedicals) in the drinking water for 5–10 days. DSS was exchanged every other day. In some experiments, mice were injected intravenously with rm-TNF (200 ng g^{−1} body weight; Immunotools) plus or minus nec-1 (1.65 µg g^{−1} body weight; Enzo). Mice were examined by measuring body temperature, weight loss and monitoring development of diarrhoea. Colitis development was monitored by analysis of rectal bleeding and high-resolution mouse video endoscopy as previously described²⁹. Mice were anaesthetized with 2–2.5% isoflurane in oxygen during endoscopy. Mice were routinely screened for pathogens according to FELASA guidelines. Animal protocols were approved by the Institutional Animal Care and Use Committee of the University of Erlangen.

Human samples. Paraffin-embedded specimens from the terminal ileum of control patients and patients with active Crohn's disease were obtained from the Institute of Pathology of the University Clinic Erlangen. The specimens had been taken from routine diagnostic samples and patient data had been made anonymous. Electron microscopy and tissue culture experiments were performed with endoscopic biopsy specimens collected in the endoscopy ward of the Department of Medicine I. The collection of samples was approved by the local ethical committee and the institutional review board of the University of Erlangen-Nuremberg and each patient gave written informed consent.

Histology, immunohistochemistry and electron microscopy. Histopathological analysis was performed on formalin-fixed paraffin-embedded tissue after H&E staining. Immunofluorescence of cryosections was performed using the TSA Cy3 system as recommended by the manufacturer (PerkinElmer). Fluorescence microscopy (Olympus) and confocal microscopy (Leica TCS SP5) was used for analysis. The following primary antibodies were used: CD4 (BD Bioscience), myeloperoxidase (Zymed Labs), F4/80 (MD Bioscience), caspase-8 (Sigma), cleaved caspase-8, cleaved caspase-3 (Cell Signaling Technology), lysozyme, chromogranin-A (Invitrogen), human RIP3 (Abcam), mouse RIP3 (AbD Serotec) and TNF-α (Pharmingen). Slides were then incubated with biotinylated secondary antibodies (Dianova). The nuclei were counterstained with Hoechst 3342 (Invitrogen). Cell death was analysed using CaspACE FITC-VAD-FMK (Promega) for early apoptosis and the *in situ* cell death detection kit (Roche) for TUNEL. For electron microscopy, glutaraldehyde-fixed material was used. After embedding in Epon Araldite, ultrathin sections were cut and analysed using a Zeiss EM 906.

Crypt isolation and organoid culture. Organ culture of freshly isolated human small intestinal biopsies was performed in RPMI medium (Gibco). For organoid culture, crypts were isolated from the small intestine of mice and cultured for a

minimum of 7 days as previously described¹¹. In brief, crypts were isolated by incubating pieces of small intestine in isolation buffer (phosphate buffered saline without calcium and magnesium (PBS0), 2 mM EDTA). Crypts were then transferred into matrigel (BD Bioscience) in 48-well plates and 350 µl culture medium (advanced DMEM/F12 (Invitrogen), containing HEPES (10 mM; PAA), GlutaMax (2 mM; Invitrogen), penicillin (100 U ml^{−1}; Gibco), streptomycin (100 µg ml^{−1}; Gibco), murine EGF (50 ng ml^{−1}; Immunotools), recombinant human R-spondin (1 µg ml^{−1}; R&D Systems), N2 Supplement 1 × (Invitrogen), B27 Supplement 1 × (Invitrogen), 1 mM N-acetylcystein (Sigma-Aldrich) and recombinant murine Noggin (100 ng ml^{−1}; Peprotech)). Organoid growth was monitored by light microscopy. In some experiments, human biopsies or organoids were treated with recombinant mouse TNF-α (25 ng ml^{−1}; Immunotools), recombinant human TNF-α (50 ng ml^{−1}; Immunotools), nec-1 (30 µM; Enzo) or caspase-8 inhibitor (50 µM; Santa Cruz). Cell viability of organoids was analysed indirectly by quantification of relative ATP level with the CellTiter-Glo assay from Promega according to the manufacturer's instructions. Luminescence was measured on the microplate reader infinite M200 (Tecan).

IEC isolation and immunoblotting. IECs were isolated in an EDTA separation solution as previously described⁵. Protein extracts were prepared using the mammalian protein extraction reagent (Thermo Scientific) supplemented with protease and phosphatase inhibitor tablets (Roche). Protein extracts were separated by SDS-PAGE (10%) and transferred to nitrocellulose transfer membranes (Whatman). Membranes were probed with the following primary antibodies: cleaved caspase-8, cleaved caspase-3, cleaved caspase-9 (Cell Signaling), RIP3 (Enzo), actin (Santa Cruz Biotechnology) and secondary HRP-linked anti-rabbit antibody (Cell Signaling).

Gene expression analyses. Total RNA was extracted from gut tissue or isolated IECs using an RNA isolation kit (Nucleo Spin RNA II, Macherey Nagel). cDNA was synthesized by reverse transcription (iScript cDNA Synthesis Kit, Bio Rad) and analysed by real-time PCR with SsoFast EvaGreen (Bio-Rad) reagent and QuantiTect Primer assays (Qiagen). Experiments were normalized to the level of the housekeeping gene *HPRT*. For gene-chip experiments total RNA of IECs from three control and three *Casp8^{ΔIEC}* mice was isolated using the RNeasy Mini Kit (Qiagen) and were performed by the Erlangen University core facility using the Affymetrix mouse 430 2.0 chip (Affymetrix). For multiple gene array testing including differential expression analysis the software package FlexArray (<http://genomequebec.mcgill.ca/FlexArray>) was used. GO-based analyses were performed using the online tool Database for Annotation, Visualization and Integrated Discovery (DAVID).

Caspase activity. Primary isolated intestinal epithelial cells were cultured with RPMI (Gibco), supplemented with 10% FCS (PAA), penicillin (100 U ml^{−1}; Gibco), streptomycin (100 µg ml^{−1}; Gibco) in fibronectin (BD Bioscience) coated 48-well plates and caspase-3/-7 activity was measured using the Caspase-Glo3/7 Assay from Promega according to the manufacturer's instructions. Luminescence was measured on the microplate reader infinite M200 (Tecan).

Statistical analysis. Data were analysed by Student's *t*-test using Microsoft Excel. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

Ebola virus entry requires the cholesterol transporter Niemann–Pick C1

Jan E. Carette^{1†*}, Matthijs Raaben^{2*}, Anthony C. Wong^{3*}, Andrew S. Herbert⁴, Gregor Obernosterer^{1†}, Nirupama Mulherkar³, Ana I. Kuehne⁴, Philip J. Kranzusch², April M. Griffin², Gordon Ruthel⁴, Paola Dal Cin⁵, John M. Dye⁴, Sean P. Whelan², Kartik Chandran³ & Thijn R. Brummelkamp^{1†}

Infections by the Ebola and Marburg filoviruses cause a rapidly fatal haemorrhagic fever in humans for which no approved antivirals are available¹. Filovirus entry is mediated by the viral spike glycoprotein (GP), which attaches viral particles to the cell surface, delivers them to endosomes and catalyses fusion between viral and endosomal membranes². Additional host factors in the endosomal compartment are probably required for viral membrane fusion; however, despite considerable efforts, these critical host factors have defied molecular identification^{3–5}. Here we describe a genome-wide haploid genetic screen in human cells to identify host factors required for Ebola virus entry. Our screen uncovered 67 mutations disrupting all six members of the homotypic fusion and vacuole protein-sorting (HOPS) multisubunit tethering complex, which is involved in the fusion of endosomes to lysosomes⁶, and 39 independent mutations that disrupt the endo/lysosomal cholesterol transporter protein Niemann–Pick C1 (NPC1)⁷. Cells defective for the HOPS complex or NPC1 function, including primary fibroblasts derived from human Niemann–Pick type C1 disease patients, are resistant to infection by Ebola virus and Marburg virus, but remain fully susceptible to a suite of unrelated viruses. We show that membrane fusion mediated by filovirus glycoproteins and viral escape from the vesicular compartment require the NPC1 protein, independent of its known function in cholesterol transport. Our findings uncover unique features of the entry pathway used by filoviruses and indicate potential antiviral strategies to combat these deadly agents.

We have developed haploid genetic screens to gain insight into the biological processes relevant to human disease^{8,9}. Here we use this approach to explore the filovirus entry pathway at an unprecedented level of detail. To interrogate millions of gene disruption events for defects in Ebola virus entry, we used a replication-competent vesicular stomatitis virus bearing the Ebola virus glycoprotein (rVSV-GP-EboV)¹⁰. Although this virus replicates in most cell lines, it inefficiently killed near-haploid KBM7 cells (Supplementary Fig. 1c). In an unsuccessful attempt to induce pluripotency in KBM7 cells by expression of OCT4 (also called POU5F1), SOX2, MYC and KLF4 (ref. 11), we obtained HAP1 cells (Supplementary Fig. 1a). HAP1 cells grew adherently and no longer expressed haematopoietic markers (Supplementary Fig. 1b). Most of these cells in early passage cultures were haploid for all chromosomes, including chromosome 8 (which is diploid in KBM7 cells). Unlike KBM7 cells, HAP1 cells were susceptible to rVSV-GP-EboV (Supplementary Fig. 1c), allowing screens for filovirus host factors.

We used a retroviral gene-trap vector⁹ to mutagenize early-passage HAP1 cells. To generate a control data set, we mapped ~800,000 insertions using deep sequencing (Supplementary Table 1). Next, we selected rVSV-GP-EboV-resistant cells, expanded them as a pool, and mapped insertion sites. Enrichment for mutations in genes was calculated by

comparing a gene's mutation frequency in resistant cells to that in the control data set (Supplementary Fig. 2). We identified a set of genes enriched for mutations in the rVSV-GP-EboV-resistant cell population (Fig. 1a, Supplementary Fig. 3 and Supplementary Table 2). Nearly all of these candidate host factors are involved in the architecture and trafficking of endo/lysosomal compartments. Our screen identified cathepsin B (CTSB), the only known host factor for which deletion inhibits Ebola virus entry⁵. Further inspection showed that mutations were highly enriched in genes encoding all six subunits of the HOPS complex (*VPS11*, *VPS16*, *VPS18*, *VPS33A*, *VPS39* and *VPS41*), for which we identified 67 independent mutations. The HOPS complex mediates fusion of endosomes and lysosomes⁶ and affects endosome maturation^{12,13}. The identification of all members of the HOPS complex demonstrates high, and possibly saturating, coverage of our screen. We also identified factors involved in the biogenesis of endosomes (PIKFYVE, FIG4)¹⁴, lysosomes (BLOC1S1, BLOC1S2)¹⁵, and in targeting of luminal cargo to the endocytic pathway (GNPTAB)¹⁶. The strongest hit was the Niemann–Pick disease locus *NPC1*, encoding an endo/lysosomal cholesterol transporter⁷. NPC1 also affects endosome/lysosome fusion and fission¹⁷, calcium homeostasis¹⁸ and HIV-1 release¹⁹.

We subcloned the resistant cell population to obtain clones deficient for *VPS11*, *VPS33A* and *NPC1* (Supplementary Fig. 4a, b and Fig. 1b). These mutants displayed marked resistance to infection by rVSV-GP-EboV and VSV pseudotyped with Ebola virus or Marburg virus GP (Fig. 1c and Supplementary Fig. 4c). Cells lacking a functional HOPS complex or NPC1 were nonetheless fully susceptible to infection by a large panel of other enveloped and non-enveloped viruses, including VSV and recombinant VSV bearing different viral glycoproteins (Fig. 1d and Supplementary Fig. 5). The susceptibility of HAP1 clones to rVSV-GP-EboV infection was restored by expression of the corresponding cDNAs (Supplementary Fig. 6a–c).

Loss of NPC1 causes Niemann–Pick disease, a neurovisceral disorder characterized by cholesterol and sphingolipid accumulation in lysosomes⁷. We tested the susceptibility of patient primary fibroblasts to filovirus-GP-dependent infection. *NPC1*-mutant cells were infected poorly or not at all by rVSV-GP-EboV and VSV pseudotyped with filovirus GP proteins (Fig. 2a, b), and infection was restored by expression of wild-type NPC1 (Fig. 2c).

Mutations in *NPC2* cause identical clinical symptoms and phenotype defects in lipid transport²⁰. Surprisingly, *NPC2*-mutant fibroblasts derived from different patients were susceptible to filovirus-GP-dependent infection (Fig. 2a, b and Supplementary Fig. 7), despite a similar accumulation of cholesterol in *NPC2*- and *NPC1*-mutant cells (Fig. 2a). Moreover, cholesterol clearance from *NPC1*-null cells by cultivation in lipoprotein-depleted growth medium did not confer susceptibility (Supplementary Fig. 8). Therefore, resistance of NPC1-deficient

¹Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA. ²Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York 10461, USA. ⁴US Army Medical Research Institute of Infectious Diseases, 1425 Porter St, Fort Detrick, Maryland 21702-5011, USA. ⁵Center for Advanced Molecular Diagnostics, Shapiro 5-058, 70 Francis Street, Boston, Massachusetts 02115, USA. [†]Present addresses: Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94304, USA (J.E.C.); Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands (G.O., T.R.B.).

*These authors contributed equally to this work.

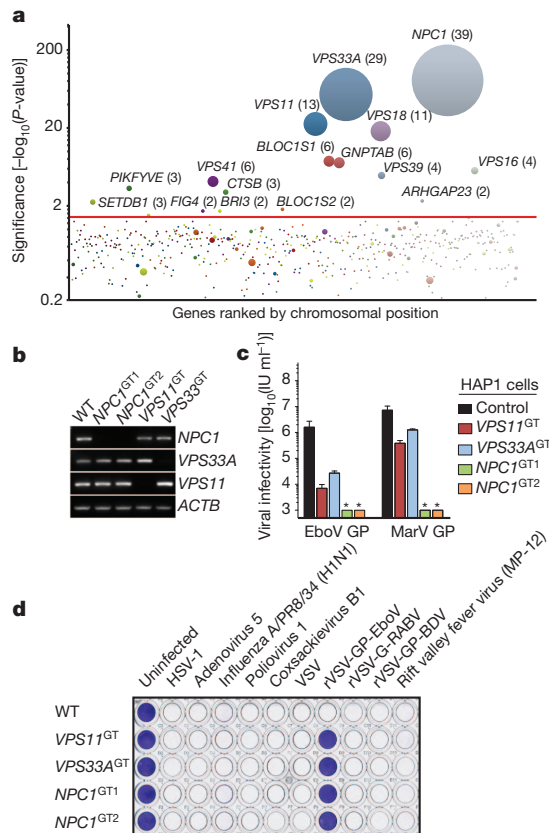


Figure 1 | A haploid genetic screen identifies the HOPS complex and NPC1 as host factors for filovirus entry. **a**, Genes enriched for gene-trap insertions in the rVSV-GP-EboV-selected cell population compared to unselected control cells. Circles represent genes and their size corresponds to the number of independent insertions identified in the rVSV-GP-EboV-selected population. Genes are ranked on the *x*-axis based on chromosomal position. **b**, RT-PCR analysis of the expression levels of *NPC1*, *VPS33A* and *VPS11* in mutant clones. **c**, Infectivity of VSV pseudotyped with the indicated filovirus glycoproteins. IU, infectious units. Means \pm standard deviation (s.d.) ($n = 3$) are shown. EboV, Ebola virus (Zaire); MarV, Marburg virus. Asterisk indicates below detection limit. **d**, HAP1 clones were infected with viruses including recombinant VSV viruses carrying rabies or Borna disease virus glycoproteins (rVSV-G-RABV and rVSV-GP-BDV) and stained with crystal violet.

cells to rVSV-GP-EboV is not caused by defects in cholesterol transport per se.

Filoviruses display broad mammalian host and tissue tropism^{21,22}. To determine if NPC1 is generally required for filovirus-GP-mediated infection, we used *Npc1*-null Chinese hamster ovary (CHO) cells. Loss of NPC1 conferred complete resistance to viral infection (Supplementary Fig. 6d) that was reversed by expression of human NPC1 (Supplementary Fig. 6e). Certain small molecules such as U18666A (ref. 23) and the antidepressant imipramine²⁴ cause a cellular phenotype similar to *NPC1* deficiency possibly by targeting NPC1 (ref. 23). Prolonged U18666A treatment has been reported to modestly inhibit VSV²⁵. However, we found that brief exposure of Vero cells and HAP1 cells to U18666A or imipramine potentially inhibited viral infection mediated by Ebola virus GP but not VSV or rabies virus G (Fig. 2d and Supplementary Figs 9 and 10). Because U18666A inhibits rVSV-GP-EboV infection only when added at early time points, it probably affects entry rather than replication (Supplementary Fig. 10). Thus, NPC1 has a critical role in infection mediated by filovirus glycoproteins that is conserved in mammals and probably independent of NPC1's role in cholesterol transport.

Filoviruses bind to one or more cell-surface molecules^{2,26,27} and are internalized by macropinocytosis^{28,29}. In *VPS33A*- and *NPC1*-mutant cells, we observed no significant differences in binding or internalization

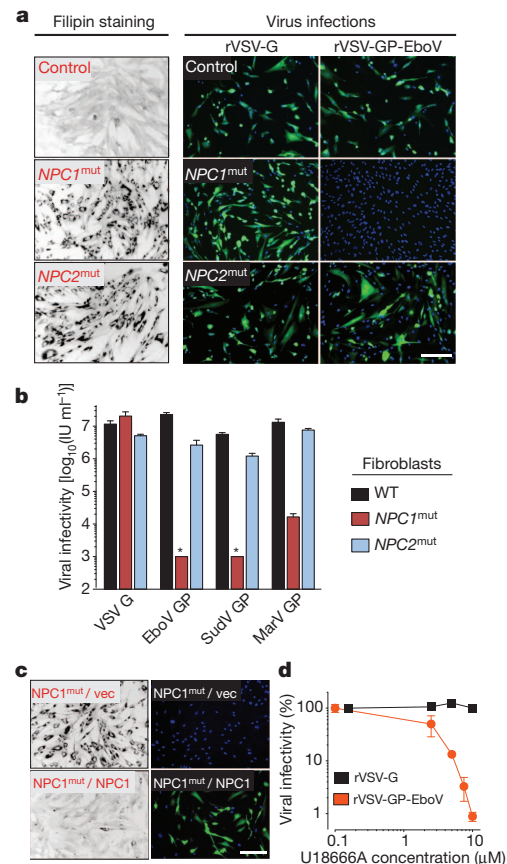


Figure 2 | Viral infection mediated by filovirus glycoproteins requires NPC1 but not NPC2. **a**, Primary skin fibroblasts from a healthy individual and patients carrying homozygous mutations in *NPC1* or *NPC2* were stained with filipin, or challenged with rVSV-G or rVSV-GP-EboV. Filipin-stained (black) and infected cells (green) were visualized by fluorescence microscopy. Filipin-stained images were inverted for clarity. Blue indicates Hoechst nuclear counterstain. **b**, Infectivity of VSV pseudotyped with the indicated viral glycoproteins in control and Niemann-Pick fibroblasts. Asterisk indicates below detection limit. SudV, Sudan virus. **c**, NPC1 patient fibroblasts expressing empty vector or human NPC1 were stained with filipin or challenged with rVSV-GP-EboV. **d**, Infectivity of rVSV-G and rVSV-GP-EboV in Vero cells pre-incubated for 30 min with the indicated concentrations of U18666A. Scale bars, 200 μ m (**a**, **c**). Means \pm s.d. ($n = 3-6$) are shown (**b**, **d**).

of Alexa-647-labelled rVSV-GP-EboV (Fig. 3a and Supplementary Figs 11 and 12a). Similar results were obtained by flow cytometry using fluorescent Ebola-virus-like particles (Supplementary Fig. 12b). Moreover, bullet-shaped VSV particles were readily observed by electron microscopy at the cell periphery and within plasma membrane invaginations resembling nascent macropinosomes (Fig. 3b). Finally, *VPS33A*- and *NPC1*-null cells were fully susceptible to vaccinia virus entry by macropinocytosis (Supplementary Fig. 13). Thus, GP-mediated entry is not inhibited at viral attachment or early internalization steps in NPC1- or HOPS-defective cells, indicating a downstream defect.

Cathepsin L (CATL; also called CTSL1)-assisted cleavage of Ebola virus GP by CTSL1 is required for viral membrane fusion^{3,5}. Mutant HAP1 cells possess normal CTSL1 activity (Supplementary Fig. 14b, c) and were fully susceptible to mammalian reoviruses, which use CTSL1 or CATL for entry (Supplementary Fig. 14d). Moreover, these cells remained refractory to *in vitro*-cleaved rVSV-GP-EboV particles (Fig. 3c) that no longer required CTSL1/CATL activity within Vero cells (Supplementary Fig. 14a). Therefore the HOPS complex and NPC1 are probably required downstream of the initial GP proteolytic processing steps that generate a primed entry intermediate.

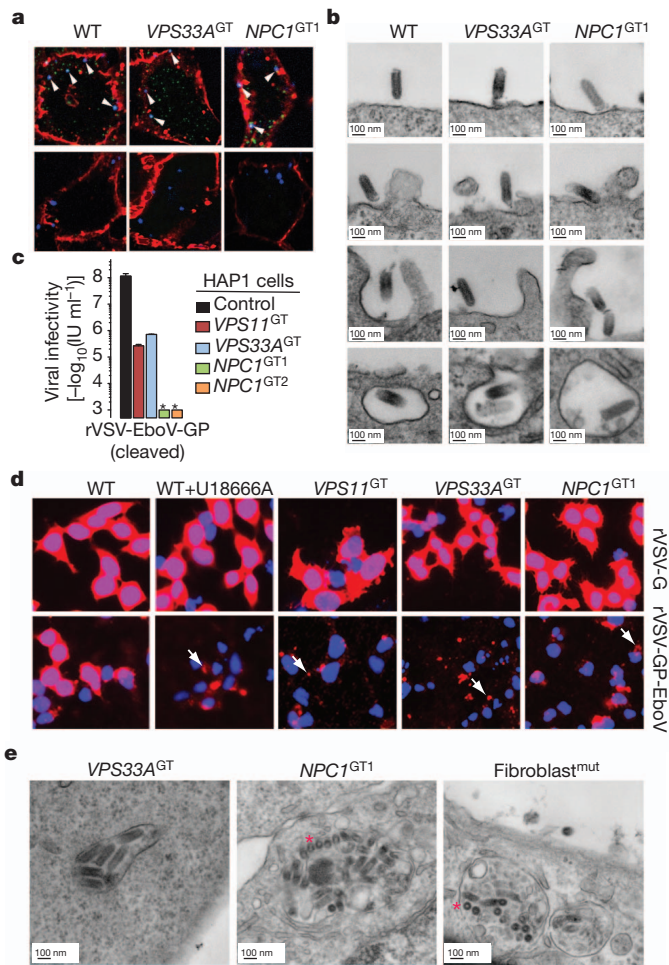


Figure 3 | Virus entry is arrested at a late step in cells deficient for the HOPS complex and NPC1. **a**, Viral particles attach and internalize into HOPS- and NPC1-deficient cells. Indicated HAP1 clones were infected with Alexa-647-labelled rVSV-GP-EboV (blue) at 4 °C. Non-internalized, bound viral particles (arrowheads, blue) were also stained with a GP-specific antibody (green) and the plasma membrane with Alexa-594-wheat germ agglutinin (red) (top panels). To assess viral internalization, cells were heated to 37 °C (bottom panels). Internalized viral particles (blue puncta) are resistant to acid-stripping and inaccessible to a GP antibody. Original magnification, $\times 63$. **b**, Cells were inoculated with rVSV-GP-EboV and examined by transmission electron microscopy. Representative images of early entry steps are shown. **c**, *In vitro*-cleaved rVSV-GP-EboV cannot bypass the infection block observed in VPS11^{GT}, VPS33A^{GT} and NPC1^{GT1} cells. GT, gene trap. Infectivity of thermolysin-cleaved rVSV-GP-EboV in the indicated HAP1 clones is shown. Asterisk indicates below the limit of detection. **d**, Viral escape into the cytoplasm is blocked in HOPS-complex- and NPC1-deficient cells. Wild-type HAP1 cells treated with U18666A (10 $\mu\text{g ml}^{-1}$) and the indicated mutant clones were infected with rVSV-G or rVSV-GP-EboV virus for 3 h and processed for VSV M staining (red). Punctate staining is indicated by arrows. Original magnification, $\times 20$. **e**, Electron micrographs of rVSV-GP-EboV-infected VPS33A- and NPC1-deficient HAP1 cells and NPC1-deficient fibroblasts showing agglomerations of bullet-shaped VSV particles in vesicular compartments. All images were taken at 3 h after inoculation. Asterisks highlight rVSV-GP-EboV particles in cross-section.

Finally, we used the intracellular distribution of the internal VSV M (matrix) protein as a marker for membrane fusion (Fig. 3d). Cells were infected with native VSV or rVSV-GP-EboV and immunostained to visualize the incoming M protein. Endosomal acid-pH-dependent entry of either virus into wild-type HAP1 cells caused redistribution of the incoming viral M throughout the cytoplasm (Fig. 3d and Supplementary Fig. 15a). By contrast, only punctate, perinuclear M staining was obtained in drug-treated and mutant cells infected with

rVSV-GP-EboV or rVSV-GP-MarV (Fig. 3d and Supplementary Fig. 15b). Electron micrographs of mutant cells infected with rVSV-GP-EboV revealed agglomerations of viral particles within vesicular compartments (Fig. 3e and Supplementary Fig. 16a) containing LAMP1 (Supplementary Fig. 16b), indicating that fusion and uncoating of incoming virus is arrested. Similarly, U18666A treatment increased the number of viral particles in NPC1- and LAMP1-positive endosomes (Supplementary Fig. 17). Therefore, NPC1 and the HOPS complex are required for late step(s) in filovirus entry leading to viral membrane fusion and escape from the lysosomal compartment.

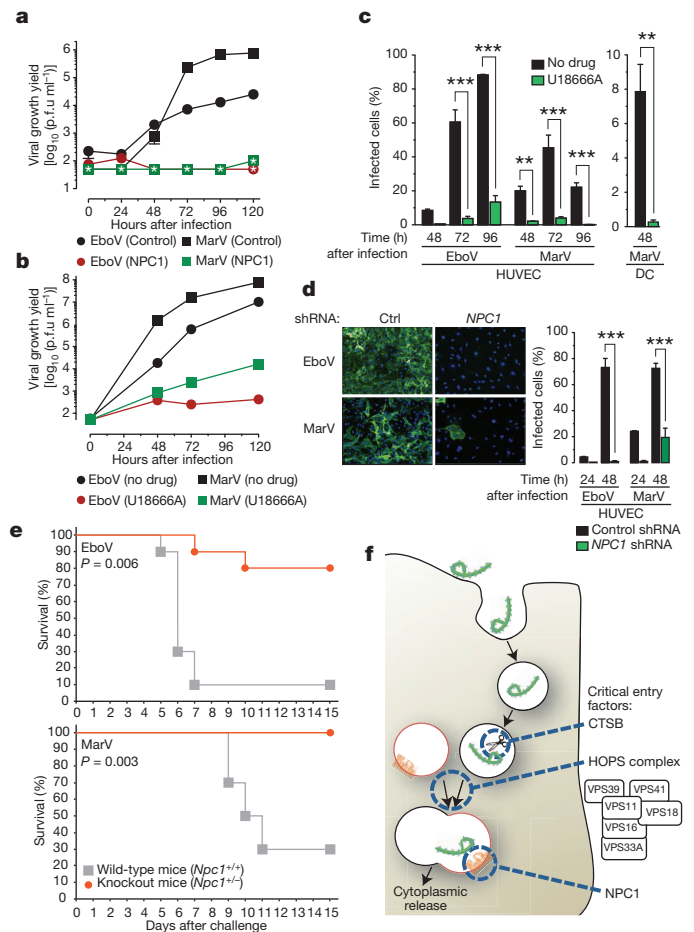


Figure 4 | NPC1 function is required for infection by authentic Ebola and Marburg viruses. **a**, NPC1 patient fibroblasts were exposed to Ebola virus (EboV) or Marburg virus (MarV) at a multiplicity of infection (MOI) of 0.1. Supernatants were harvested and yields of infectious virus were measured. Asterisk indicates below detection limit. p.f.u., plaque-forming units. **b**, Vero cells treated with DMSO or U18666A (20 μM) were infected with Ebola virus or Marburg virus at a MOI of 0.1 and yields of infectious virus were measured. **c**, Human peripheral blood monocyte-derived dendritic cells (DC) and umbilical-vein endothelial cells (HUVEC) were infected in the presence or absence of U18666A at a MOI of 3 and the percentage of infected cells was determined by immunostaining. **d**, HUVECs were transduced with lentiviral vectors expressing a non-targeting short hairpin (sh)RNA (Ctrl) or an shRNA targeting NPC1, infected with Ebola virus or Marburg virus at a MOI of 3 and the percentage of infected cells was determined. Representative images of cells 48 h after infection are also shown: green, viral antigen; blue, nuclear counterstain. For panels **a**–**d**, Means \pm s.d. are shown ($n = 2$ – 3). In panels **a**, **b**, error bars are not visible because they are within the symbols. For panels **c**, **d**, $**P < 0.01$; $***P < 0.001$. **e**, Survival of *Npc1*^{+/+} and *Npc1*^{+/−} mice ($n = 10$ for each group) inoculated intraperitoneally with $\sim 1,000$ p.f.u. of mouse-adapted Ebola virus or Marburg virus. **f**, A proposed hypothetical model for the roles of CTSB, the HOPS complex and NPC1 in Ebola virus entry.

We next tested if infection by authentic Ebola virus and Marburg virus is affected in *NPC1*-mutant primary patient fibroblasts. Yields of viral progeny were profoundly reduced for both viruses in mutant cells (Fig. 4a). Marked reductions in viral yield were also obtained in Vero cells treated with U18666A (Fig. 4b). Moreover U18666A greatly reduced infection of human peripheral blood monocyte-derived dendritic cells and umbilical-vein endothelial cells (HUVECs) (Fig. 4c), without affecting cell number or morphology (Supplementary Fig. 19). Finally, knock-down of *NPC1* in HUVECs diminished infection by filoviruses (Fig. 4d and Supplementary Fig. 18). These findings indicate that *NPC1* is critical for authentic filovirus infection.

We assessed the effect of *NPC1* mutation in lethal mouse models of Ebola virus and Marburg virus infection. Heterozygous *Npc1* (*Npc1*^{+/-}) knockout mice and their wild-type littermates were challenged with mouse-adapted Ebola virus or Marburg virus and monitored for 28 days. Whereas *Npc1*^{+/+} mice rapidly succumbed to infection with either filovirus, *Npc1*^{+/-} mice were largely protected (Fig. 4e).

We have used global gene disruption in human cells to discover components of the unusual entry pathway used by filoviruses. Most of the identified genes affect aspects of lysosome function, indicating that filoviruses exploit this organelle differently from all other viruses that we have tested (Fig. 4f). The unanticipated role for the hereditary disease gene *NPC1* in viral entry, infection and pathogenesis may facilitate the development of antifilovirus therapeutics.

METHODS SUMMARY

Adherent HAP1 cells were generated by the introduction of OCT4/SOX2/Myc and KLF4 transcription factors. 100 million cells were mutagenized using a retroviral gene-trap vector. Insertion sites were mapped for approximately 1% of the unselected population using parallel sequencing. Cells were infected with rVSV-GP-EboV and the resistant cell population was expanded. Genes that were statistically enriched for mutation events in the selected population were identified, and the roles of selected genes in filovirus entry were characterized.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 December 2010; accepted 30 June 2011.

Published online 24 August 2011.

- Feldmann, H. & Geisbert, T. W. Ebola haemorrhagic fever. *Lancet* **377**, 849–862 (2010).
- Lee, J. E. & Saphire, E. O. Ebolavirus glycoprotein structure and mechanism of entry. *Future Virol.* **4**, 621–635 (2009).
- Schornerberg, K. *et al.* Role of endosomal cathepsins in entry mediated by the Ebola virus glycoprotein. *J. Virol.* **80**, 4174–4178 (2006).
- Kuhn, J. H. *et al.* Conserved receptor-binding domains of Lake Victoria marburgvirus and Zaire ebolavirus bind a common receptor. *J. Biol. Chem.* **281**, 15951–15958 (2006).
- Chandran, K., Sullivan, N. J., Felbor, U., Whelan, S. P. & Cunningham, J. M. Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection. *Science* **308**, 1643–1645 (2005).
- Nickerson, D. P., Brett, C. L. & Merz, A. J. Vps-C complexes: gatekeepers of endolysosomal traffic. *Curr. Opin. Cell Biol.* **21**, 543–551 (2009).
- Carstea, E. D. *et al.* Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. *Science* **277**, 228–231 (1997).
- Carette, J. E. *et al.* Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nature Biotechnol.* **29**, 542–546 (2011).
- Carette, J. E. *et al.* Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326**, 1231–1235 (2009).
- Wong, A. C., Sandesara, R. G., Mulherkar, N., Whelan, S. P. & Chandran, K. A forward genetic strategy reveals destabilizing mutations in the Ebolavirus glycoprotein that alter its protease dependence during cell entry. *J. Virol.* **84**, 163–175 (2010).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Poteryaev, D., Datta, S., Ackema, K., Zerial, M. & Spang, A. Identification of the switch in early-to-late endosome transition. *Cell* **141**, 497–508 (2010).
- Rink, J., Ghigo, E., Kalaidzidis, Y. & Zerial, M. Rab conversion as a mechanism of progression from early to late endosomes. *Cell* **122**, 735–749 (2005).
- Sbrissa, D. *et al.* Core protein machinery for mammalian phosphatidylinositol 3,5-bisphosphate synthesis and turnover that regulates the progression of endosomal transport. Novel Sac phosphatase joins the ArPIKfyve-PIKfyve complex. *J. Biol. Chem.* **282**, 23878–23891 (2007).
- Dell'Angelica, E. C. The building BLOC(k)s of lysosomes and related organelles. *Curr. Opin. Cell Biol.* **16**, 458–464 (2004).
- Tiede, S. *et al.* Mucopolidosis II is caused by mutations in GNPTA encoding the alpha/beta GlcNAc-1-phosphotransferase. *Nature Med.* **11**, 1109–1112 (2005).
- Goldman, S. D. & Krise, J. P. Niemann-Pick C1 functions independently of Niemann-Pick C2 in the initial stage of retrograde transport of membrane-impermeable lysosomal cargo. *J. Biol. Chem.* **285**, 4983–4994 (2010).
- Lloyd-Evans, E. *et al.* Niemann-Pick disease type C1 is a sphingosine storage disease that causes deregulation of lysosomal calcium. *Nature Med.* **14**, 1247–1255 (2008).
- Tang, Y. Y., Leao, I. C., Coleman, E. M., Broughton, R. S. & Hildreth, J. E. K. Deficiency of Niemann-Pick type C-1 protein impairs release of human immunodeficiency virus type 1 and results in Gag accumulation in late endosomal/lysosomal compartments. *J. Virol.* **83**, 7982–7995 (2009).
- Naureckiene, S. *et al.* Identification of HE1 as the second gene of Niemann-Pick C disease. *Science* **290**, 2298–2301 (2000).
- Takada, A. *et al.* A system for functional analysis of Ebola virus glycoprotein. *Proc. Natl Acad. Sci. USA* **94**, 14764–14769 (1997).
- Wool-Lewis, R. J. & Bates, P. Characterization of Ebola virus entry by using pseudotyped viruses: identification of receptor-deficient cell lines. *J. Virol.* **72**, 3155–3160 (1998).
- Cenedella, R. J. Cholesterol synthesis inhibitor U18666A and the role of sterol metabolism and trafficking in numerous pathophysiological processes. *Lipids* **44**, 477–487 (2009).
- Rodriguez-Lafrasse, C. *et al.* Abnormal cholesterol metabolism in imipramine-treated fibroblast cultures. Similarities with Niemann-Pick type C disease. *Biochim Biophys Acta* **1043**, 123–128 (1990).
- Sobo, K. *et al.* Late endosomal cholesterol accumulation leads to impaired intra-endosomal trafficking. *Plos One* **2**, e851 (2007).
- Kondratowicz, A. S. *et al.* T-cell immunoglobulin and mucin domain 1 (TIM-1) is a receptor for Zaire Ebolavirus and Lake Victoria Marburgvirus. *Proc. Natl Acad. Sci. USA* **108**, 8426–8431 (2011).
- Alvarez, C. P. *et al.* C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus *in cis* and *in trans*. *J. Virol.* **76**, 6841–6844 (2002).
- Saeed, M. F., Kolokoltsov, A. A., Albrecht, T. & Davey, R. A. Cellular entry of ebola virus involves uptake by a macropinocytosis-like mechanism and subsequent trafficking through early and late endosomes. *PLoS Pathog.* **6**, <http://dx.doi.org/10.1371/journal.ppat.1001110> (2010).
- Nanbo, A. *et al.* Ebolavirus is internalized into host cells via macropinocytosis in a viral glycoprotein-dependent manner. *PLoS Pathog.* **6**, <http://dx.doi.org/10.1371/journal.ppat.1001121> (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank M. Kielian, H. Ploegh, V. Prasad and D. Sabatini for critical reading of the manuscript and valuable advice; C. Guimaraes, V. Blomen and T. Peterson for suggestions; M. Bogoy for providing the CTSS/CATL activity probe (GB111); T.-Y. Chang for the gift of NPC1-null CHO cells; D. Lyles for the antibody to VSV M; M. Nibert for providing reovirus; J. de la Torre for providing rVSV-GP-BDV; J. Wojcechowskyj for providing RVF; E. Mühlberger for providing Ebola cDNA; and M. Ericsson for support with electron microscopy. This research was supported by NIH grants R01 AI088027 (K.C.), AI081842 and U54 AI057159 (NERCE-BEID) (S.P.W.), and R21 HG004938 (T.R.B.), and by the DTRA Project, CBM.VAXPLAT.05.10.RD.005 (J.M.D.). T.R.B. was additionally supported by the Whitehead Fellows Program. S.P.W. is a recipient of a Burroughs Wellcome Investigators in the Pathogenesis of Infectious Disease Award. A.C.W. was additionally supported by NIH-funded training programs T32 GM007288 and T32 AI070117 at the Albert Einstein College of Medicine. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the US Army.

Author Contributions K.C., S.P.W., T.R.B. and J.M.D. were the senior authors of this study and made equivalent contributions. The study was conceived by K.C., S.P.W. and T.R.B. J.E.C. and T.R.B. devised and implemented the haploid genetic screen, generated the HAP1 cells and identified hits by deep sequencing and cell cloning. P.D.C. carried out karyotype analysis on the HAP1 line. K.C. created and characterized the rVSV-GP-EboV virus used in the screen. A.M.G. created the rVSV-G-RABV. J.E.C., G.O. and K.C. performed entry and infection experiments with the HAP1 cells. A.C.W. and K.C. carried out entry and infection experiments with rVSVs in human fibroblasts, CHO and Vero cells. N.M. and K.C. carried out RNAi experiments with primary cells. M.R. was involved in experimental strategy and design and performed entry and infection experiments by high-resolution fluorescence and electron microscopy. N.M. carried out VLP entry experiments and P.J.K., the replicon assay. A.C.W. performed the cysteine cathepsin enzyme assays. A.S.H., A.I.K. and J.M.D. performed the infection and animal challenge experiments with the authentic viral agents. G.R. performed fluorescence microscopy and image analysis with filovirus-infected cell cultures. J.E.C., K.C., S.P.W. and T.R.B. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.R.B. (tbrummelkamp@nki.nl), K.C. (kartik.chandran@einstein.yu.edu), S.P.W. (sean_whelan@hms.harvard.edu) or J.M.D. (John.M.Dye1@us.army.mil).

METHODS

Cells. KBM7 cells and derivatives were maintained in IMDM supplemented with 10% FCS, L-glutamine, and penicillin–streptomycin. Vero cells and primary human dermal fibroblasts (Coriell Institute for Medical Research) were maintained in DMEM supplemented with 10% FCS, L-glutamine and penicillin–streptomycin. Wild-type and NPC1-null (CT43) Chinese hamster ovary (CHO) fibroblasts were maintained in DMEM–Ham's F-12 medium (50–50 mix) supplemented with 10% FCS, L-glutamine and penicillin–streptomycin³⁰.

To generate dendritic cells, primary human monocytes were cultured at 37 °C, 5% CO₂, and 80% humidity in RPMI supplemented with 10% human serum, L-glutamine, sodium pyruvate, HEPES, penicillin–streptomycin, recombinant human granulocyte monocyte-colony stimulating factor (50 ng ml⁻¹) and recombinant human interleukin-4 (50 ng ml⁻¹) for 6 days. Cytokines were added every 2 days by replacing half of the culture volume with fresh culture media. Dendritic cells were collected on day 6, characterized by flow cytometry (see below) and used immediately. Human umbilical vein endothelial cells (HUVECs) were obtained from Lonza and maintained in endothelial growth medium (EGM; Lonza).

HAP1 cells were used for the haploid screen and fibroblasts or CHO cells were used for hit validation and functional studies. Vero cells are commonly used in studies of filovirus replication, because they are highly susceptible to infection. Dendritic cell and HUVECs resemble cell types that are early and late targets of filovirus infection *in vivo*, respectively^{31,32}.

Flow cytometry of dendritic cells. Human dendritic cells were treated with Fc-block (BD Pharmingen) before incubation with mouse anti-human CD11c-APC (BioLegend) and mouse anti-human CD209-PE or isotype controls. Dendritic cells were washed and re-suspended in PBS for flow cytometric analysis using a BD FACSCanto II flow cytometer (BD Biosciences). Data analysis was completed using FlowJo software. >95% of cells were routinely observed to be CD11c⁺, DC-SIGN⁺.

Viruses. Recombinant VSV expressing eGFP and Ebola virus GP (rVSV-GP-EboV) was recovered and amplified as described¹⁰. Recombinant rVSV-GP-BDV was provided by J. C. de la Torre. rVSV-G-RABV was generated by replacement of the VSV G ORF in VSV-eGFP³³ with that of the SAD-B19 strain of rabies virus, and recombinant virus was recovered and amplified³⁴. VSV pseudotypes bearing glycoproteins derived from Ebola virus, Sudan virus and Marburg virus were generated as described³⁵.

The following non-recombinant viruses were used: adenovirus type 5 (ATCC), coxsackievirus B1 (ATCC), poliovirus 1 Mahoney (provided by C. Schliker), HSV-1 KOS (provided by H. Ploegh), influenza A/PR8/34 (H1N1) (Charles Rivers), Rift valley fever virus MP-12 (provided by J. Wojcechowskyj), and mammalian reovirus serotype 1 (provided by M. Nibert).

Generation of HAP1 cells. Retroviruses encoding SOX2, MYC, OCT4 and KLF4 were produced³⁶. Concentrated virus was used to infect near-haploid KBM7 cells in three consecutive rounds of spin-infection with an interval of 12 h. Colonies were picked and tested for ploidy. One clonally derived cell line (referred to as HAP1) was further grown and characterized. Karyotyping analysis demonstrated that most cells (27 of 39) were fully haploid, a smaller population (9 of 39) was haploid for all chromosomes except chromosome 8, like the parental KBM7 cells. Less than 10% (3 of 39) was diploid for all chromosomes except for chromosome 8, which was tetraploid.

Haploid genetic screen. Gene-trap virus was produced in 293T cells by transfection of pGT-GFP, pGT-GFP+1 and pGT-GFP+2 combined with pAdvantage, CMV-VSVG and Gag-pol. The virus was concentrated using ultracentrifugation for 1.5 h at 25,000 r.p.m. in a Beckman SW28 rotor. 100 million HAP1 cells were infected. A proportion of the cells was harvested for genomic DNA isolation to create a control data set. For the screen, 100 million mutagenized cells were exposed to rVSV-GP-EboV at a MOI ~100. The resistant colonies were expanded and ~30 million cells were used for genomic DNA isolation.

Sequence analysis of gene-trap insertion sites. Insertion sites were identified by sequencing the genomic DNA flanking gene-trap proviral DNA as described before⁸. In short, a control data set was generated containing insertion sites in mutagenized HAP1 cells before selection with rVSV-GP-EboV. Genomic DNA was isolated from ~40 million cells and subjected to a linear PCR followed by linker ligation, PCR and sequencing using the Genome Analyser platform (Illumina). Insertion sites were mapped to the human genome and insertion sites were identified that were located in Refseq genes. Insertions in this control data set comprise ~400,000 independent insertions that meet this criteria (Supplementary Table 1). To generate the experimental data set, insertions in the mutagenized HAP1 cells after selection with rVSV-GP-EboV were identified using an inverse PCR protocol followed by sequencing using the Genome Analyser. The number of inactivating mutations (that is, sense orientation or present in exon) per individual gene was counted as well as the total number of inactivating insertions for all genes. Enrichment of a gene in the screen was calculated by comparing how often that

gene was mutated in the screen compared to how often the gene carries an insertion in the control data set. For each gene a *P*-value (corrected for false discovery rate) was calculated using the one-sided Fisher exact test (Supplementary Table 2).

Characterization of the HAP1 mutant lines. Genomic DNA was isolated using Qiamp DNA mini kit (Qiagen). To confirm that the cells were truly clonal and to confirm the absence of the wild-type DNA locus, a PCR was performed with primers flanking the insertion site using the following primers: (NPC-F1, 5'-GAAGTTGGTCTGGCGATGGAG-3'; NPC1-R2, 5'-AAGGTCCTGATCTAACTCTAG-3'; VPS33A-F1, 5'-TGTCCTACGGCCGAGTGAACC-3'; VPS33A-R1, 5'-CTGTACACTTTGCTCAGTTTC-3'; VPS11-F1, 5'-GAAGGAGCCGTGAGCAATGATG-3'; VPS11-R1, 5'-GGCCAGAATTTAGTAGCAGCACT-3'). To confirm the correct insertion of the gene trap at the different loci a PCR was performed using the reverse (R1) primers of NPC1, VPS11 and VPS33A combined with a primer specific for the gene trap vector: PGT-F1; 5'-TCTCCAAATCTCGGTGGAAC-3'. To determine RNA expression levels of NPC1, VPS11 and VPS33A, total RNA was reverse transcribed using Superscript III (Invitrogen) and amplified using gene-specific primers: VPS11, 5'-CTGCTTCAAGTTCCTTTGCG-3' and 5'-AAGATTCGAGTGCAGAGTGG-3'; NPC1, 5'-CCACAGCATGACCGCTC-3' and 5'-CAGCTCACAAAACAGGTTTCAG-3'; VPS33A, 5'-TTAACACCTCTTGCCACTCAG-3' and 5'-TGTGTCTTTCTCTGCAATGCTG-3'.

Generation of stable cell populations expressing an NPC1-Flag fusion protein. A human cDNA encoding NPC1 (Origene) was ligated in-frame to a triple Flag sequence and the resulting gene encoding a C-terminally Flag-tagged NPC1 protein was subcloned into the pBABE-puro retroviral vector³⁷. Retroviral particles packaging the NPC1-Flag gene or no insert were generated by triple transfection in 293T cells, and used to infect control and NPC1-deficient human fibroblasts and CHO lines. Puromycin-resistant stable cell populations were generated.

Cell viability assays for virus treatments. KBM7 and HAP1 cells were seeded at 10,000 cells per well in 96-well tissue culture plates and treated with the indicated concentrations of rVSV-GP-EboV. After 3 days cell viability was measured using an XTT colorimetric assay (Roche). Viability is plotted as percentage viability compared to untreated control. To compare susceptibility of the HAP1 mutants to different viruses, they were seeded at 10,000 cells per well and treated with different cytolytic viruses at a concentration that in pilot experiments was the lowest concentration to produce extensive cytopathic effects. Three days after treatment, viable, adherent cells were fixed with 4% formaldehyde in phosphate-buffered saline (PBS) and stained with crystal violet.

VSV infectivity measurements. Infectivities of VSV pseudotypes were measured by manual counting of eGFP-positive cells using fluorescence microscopy at 16–26 h after infection, as described previously⁵. rVSV-GP-EboV infectivity was measured by fluorescent-focus assay (FFA), as described previously¹⁰.

Filipin staining. Filipin staining to visualize intracellular cholesterol was done as described³⁸. Cells were fixed with paraformaldehyde (3%) for 15 min at 25 °C. After three PBS washes, cells were incubated with filipin complex from *Streptomyces filipinensis* (Sigma-Aldrich) (50 µg ml⁻¹) in the dark for 1 h at room temperature. After three PBS washes, cells were visualized by fluorescence microscopy in the DAPI channel.

Measurements of cysteine cathepsin activity. Enzymatic activities of CTSB and CATL in acidified post-nuclear extracts of Vero cells, human fibroblasts and CHO lines were assayed with fluorogenic peptide substrates Z-Arg-Arg-AMC (Bachem Inc.) and (Z-Phe-Arg)-2-R110 (Invitrogen) as described³⁹. As a control for assay specificity, enzyme activities were also assessed in extracts pre-treated with E-64 (10 µM), a broad-spectrum cysteine protease inhibitor, as previously described¹⁰. Active CTSB and CATL within intact cells were labelled with the fluorescently labelled activity-based probe GB111 (1 µM) and visualized by gel electrophoresis and fluorimaging, as described previously⁴⁰.

Purification and dye conjugation of rVSV-GP-EboV. rVSV-GP-EboV was purified and labelled with Alexa Fluor 647 (Molecular Probes, Invitrogen Corporation) as described⁴¹ with minor modifications. Briefly, Alexa Fluor 647 (Molecular Probes, Invitrogen Corporation) was solubilized in DMSO at 10 mg ml⁻¹ and incubated at a concentration of 31.25 µg ml⁻¹ with purified rVSV-GP-EboV (0.5 mg ml⁻¹) in 0.1 M NaHCO₃ (pH 8.3) for 90 min at room temperature. Virus was separated from free dye by ultracentrifugation. Labelled viruses were re-suspended in NTE (10 mM Tris pH 7.4, 100 mM NaCl, 1 mM EDTA) and stored at –80 °C.

Virus binding/internalization assay. Cells were inoculated with a MOI of 200–500 of Alexa-647-labelled rVSV-GP-EboV at 4 °C for 30 min to allow binding of virus to the cell surface. Cells were subsequently fixed in 2% paraformaldehyde (to examine virus binding) or after a 2-h incubation at 37 °C and an acid wash to remove surface-bound virus. The cellular plasma membrane was labelled by incubation of

cells with $1 \mu\text{g ml}^{-1}$ Alexa Fluor 594 wheat germ agglutinin (Molecular Probes, Invitrogen) in PBS for 15 min at room temperature. External virus particles were detected using a 1:2,000 dilution of antibody 265.1, a mouse monoclonal antibody specific for Ebola GP. The GP antibodies were detected by Alexa-488-conjugated goat anti-mouse secondary antibody (Molecular Probes, Invitrogen). After washing with PBS, cells were mounted onto glass slides using Prolong Antifade Reagent (Invitrogen, Molecular Probes). Fluorescence was monitored with an epifluorescence microscope (Axiovert 200M; Carl Zeiss) equipped with a $\times 63$ objective and representative images were acquired using Slidebook 4.2 software (Intelligent Imaging Innovations)^{41,42}.

VSV M protein-release assay. Cells grown on 12-mm coverslips coated with poly-D-lysine (Sigma-Aldrich) were pre-treated with $5 \mu\text{g ml}^{-1}$ puromycin for 30 min and inoculated with rVSV at a MOI of 200–500 in the presence of puromycin. After 3 h, cells were washed once with PBS and fixed with 2% paraformaldehyde in PBS for 15 min at room temperature. To detect VSV M protein, fixed cells were incubated with a 1:7,500 dilution of monoclonal antibody 23H12 (gift of D. Lyles⁴³) in PBS containing 1% BSA and 0.1% Triton X-100 for 30 min at room temperature. Cells were washed three times with PBS, and the anti-M antibodies were detected using a 1:750 dilution of Alexa 594-conjugated goat anti-mouse secondary antibodies. Cells were counter-stained with DAPI to visualize nuclei. Cells were washed three times and mounted onto glass slides after which M localization images were acquired using a Nikon TE2000-U inverted epifluorescence microscope (Nikon Instruments) equipped with a $\times 20$ objective. Representative images were acquired with Metamorph software (Molecular Devices).

Electron microscopy. Confluent cell monolayers in 6-well plates were inoculated with rVSV-GP-EboV at a MOI of 200–500 for 3 h. Cells were fixed for at least 1 h at room temperature in a mixture of 2.5% glutaraldehyde, 1.25% paraformaldehyde and 0.03% picric acid in 0.1 M sodium cacodylate buffer (pH 7.4). Samples were washed extensively in 0.1 M sodium cacodylate buffer (pH 7.4) and treated with 1% osmium tetroxide and 1.5% potassiumferrocyanide in water for 30 min at room temperature. Treated samples were washed in water, stained in 1% aqueous uranyl acetate for 30 min, and dehydrated in grades of alcohol (70%, 90%, $2\times 100\%$) for 5 min each. Cells were removed from the dish with propyleneoxide and pelleted at 3,000 r.p.m. for 3 min. Samples were infiltrated with Epon mixed with propyleneoxide (1:1) for 2 h at room temperature. Samples were embedded in fresh Epon and left to polymerize for 24–48 h at 65°C . Ultrathin sections (about 60–80 nm) were cut on a Reichert Ultracut-S microtome and placed onto copper grids. For preparation of cryosections the virus-inoculated cells were rinsed once with PBS and removed from the dish with 0.5 mM EDTA in PBS. The cell suspension was layered on top of an 8% paraformaldehyde cushion in an Eppendorf tube and pelleted for 3 min at 3,000 r.p.m. The supernatant was removed and fresh 4% paraformaldehyde was added. After 2 h incubation, the fixative was replaced with PBS. Before freezing in liquid nitrogen the cell pellets were infiltrated with 2.3 M sucrose in PBS for 15 min. Frozen samples were sectioned at -120°C and transferred to formvar-carbon-coated copper grids. Grids were stained for lysosomes with a mouse monoclonal antibody raised against LAMP1 (HA43; Santa Cruz Biotechnology). The LAMP1 antibodies were visualized with Protein-A gold secondary antibodies. Contrasting/embedding of the labelled grids was carried out on ice in 0.3% uranyl acetate in 2% methyl cellulose. All grids were examined in a TecnaiG² Spirit BioTWIN mission electron microscope and images were recorded with an AMT 2k CCD camera.

Authentic filoviruses and infections. Vero cells were pre-treated with culture medium lacking or containing U18666A (20 μM) for 1 h at 37°C . VERO cells and primary human dermal fibroblasts were exposed to Ebola virus Zaire 1995 or Marburg virus Ci67 at a MOI of 0.1 for 1 h. Viral inoculum was removed and fresh culture media with or without drug was added. Samples of culture supernatants were collected and stored at -80°C until plaque assays were completed.

Dendritic cells were collected and seeded in 96-well poly-D-lysine-coated black plates (Greiner Bio-One) at 5×10^4 cells per well or in 6-well plates at 10^6 cells per well in culture media and incubated overnight at 37°C . They were pre-treated with medium lacking or containing U18666A as described above. Dendritic cells were exposed to Ebola virus Zaire 1995 or Marburg virus Ci67 at a MOI of 3 for 1 h. Virus inoculum was removed and fresh culture media with or without drug was added. Uninfected cells with or without drug served as negative controls. Cells were incubated at 37°C and fixed with 10% formalin at designated times. HUVECs were seeded in 96-well poly-D-lysine-coated black plates at 5×10^4 cells per well in culture media, treated with U18666A, infected, and processed as described above for dendritic cells.

Cytotoxicity analysis. Dendritic cells and HUVECs were seeded in 96-well plates. After overnight incubation at 37°C , U18666A was added at the same concentrations used for the viral infection studies. Cells in culture media without drug served as the untreated control. At indicated times after treatment, an equal volume of

CellTiter-Glo Reagent (Promega) was added to wells containing cells in culture media. Luminescence was measured using a plate reader.

Plaque assays for titration of filoviruses. Tenfold serial dilutions of culture supernatants or serum were prepared in modified Eagle's medium with Earle's balanced salts and nonessential amino acids (EMEM/NEAA) plus 5% heat-inactivated fetal bovine serum. Each dilution was inoculated into a well of a 6-well plate containing confluent monolayers of Vero 76 cells. After adsorption for 1 h at 37°C , monolayers were overlaid with a mixture of 1 part of 1% agarose (Seakem) and 1 part of $2\times$ Eagle basal medium (EBME), 30 mM HEPES buffer and 5% heat-inactivated fetal bovine serum. After incubation at 37°C , 5% CO_2 , 80% humidity for 6 days, a second overlay with 5% Neutral red was added. Plaques were counted the following day, and titres were expressed as p.f.u. ml^{-1} .

Analysis of filovirus-infected cultures by immunofluorescence. Formalin-fixed cells were blocked with 1% bovine serum albumin solution before incubation with primary antibodies. Ebola-virus-infected cells and uninfected controls were incubated with Ebola virus GP-specific monoclonal antibodies 13F6 (ref. 44) or KZ52 (ref. 45). Marburg-virus-infected cells and uninfected controls were incubated with Marburg virus GP-specific monoclonal antibody 9G4. Cells were washed with PBS before incubation with either goat anti-mouse IgG or goat anti-human IgG conjugated to Alexa 488. Cells were counterstained with Hoechst stain (Molecular Probes), washed with PBS and stored at 4°C .

Image analysis. Images were acquired at 9 fields per well with a $\times 10$ objective lens on a Discovery-1 high content imager (Molecular Devices) or at 6 fields per well with a $\times 20$ objective lens on an Operetta (Perkin Elmer) high content device. Discovery-1 images were analysed with the 'live/dead' module in MetaXpress software. Operetta images were analysed with a customized scheme built from image analysis functions present in Harmony software.

Animals and filovirus challenge experiments. Mouse-adapted Ebola virus has been described⁴⁶. Mouse-adapted Marburg virus Ci67 was provided by S. Bavari⁴⁷. Female and male BALB/c *Npc1*^{+/-} mice and BALB/c *Npc1*^{+/+} mice (5–8-week-old) were obtained from Jackson Laboratory. Mice were housed under specific pathogen-free conditions. Research was conducted in compliance with the Animal Welfare Act and other federal statutes and regulations relating to animals and experiments involving animals and adhered to principles stated in the Guide for the Care and Use of Laboratory Animals (National Research Council, 1996). The facility where this research was conducted is fully accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care International. For infection, mice were inoculated intraperitoneally with a target dose of 1,000 p.f.u. ($30,000\times$ the 50% lethal dose) of mouse-adapted Ebola virus or mouse-adapted Marburg Ci67 virus in a biosafety level 4 laboratory. Mice were observed for 28 days after challenge by study personnel and by an impartial third party. Daily observations included evaluation of mice for clinical symptoms such as reduced grooming, ruffled fur, hunched posture, subdued response to stimulation, nasal discharge and bleeding. Serum was collected from surviving mice to confirm virus clearance. Back titration of the challenge dose by plaque assay determined that Ebola-virus-infected mice received 900 p.f.u. per mouse and Marburg-virus-infected mice received 700 p.f.u. per mouse.

RNA interference. Lentiviral vectors expressing an shRNA specific for NPC1 (Sigma-Aldrich; clone# TRCN0000005428; sequence CCACAAGTTCTATAC CATATT) or a non-targeting control shRNA (Sigma-Aldrich; SHC002; sequence CAACAAGATGAAGAGACACCAA) were packaged into HIV-1 pseudotype virus by transfection in HEK 293T cells and lentivirus-containing supernatants were harvested at 36 h and 48 h after transfection and centrifuged onto HUVECs in 12-well plates in the presence of $6 \mu\text{g ml}^{-1}$ polybrene at 2,500 r.p.m., 25°C for 90 min. HepG2 cells were transduced as above but without the centrifugation step. Cells were subjected to puromycin selection 24 h after the last lentiviral transduction (HepG2, $1 \mu\text{g ml}^{-1}$; HUVECs, $1.5 \mu\text{g ml}^{-1}$) for 48–72 h before harvest for experiments. The level of NPC1 knockdown was assessed by SDS-polyacrylamide gel electrophoresis of cell extracts and immunoblotting with an anti-NPC1 polyclonal antibody (Abcam).

Ebola virus replicon assay. Ebola virus support plasmids were created by cloning the NP, VP35, VP30 and L genes from cDNA (provided by E. Mühlberger⁴⁸) into pGEM3 (Promega) and the mutant pL-D742A plasmid was generated by QuikChange site-directed mutagenesis (Stratagene). Truncated versions of the Ebola virus non-coding sequence were generated by overlap-extension PCR and appended to the *eGFP* ORF. The replicon pZEm was prepared as described previously⁴⁹. The replicon RNA sequence is flanked on the 5' end by a truncated T7 promoter with a single guanosine nucleotide and on the 3' end by the HDV ribozyme sequence and T7 terminator. The transcribed replicon RNA consists of the following EboV Zaire sequences (GenBank accession AF086833): [5']-single guanosine nucleotide-176-nucleotide genomic 5' terminus-55-nucleotide L mRNA 3' UTR-*eGFP* ORF (antisense orientation)-100-nucleotide NP mRNA 5' UTR-155-nucleotide genomic 3' terminus-[3']. The viral replicon

assay was performed as described previously⁴⁹ except that U18666A (20 $\mu\text{g ml}^{-1}$) was included in the supplemented DMEM where indicated. Images were collected directly from 6-cm dishes with a Zeiss Axioplan inverted fluorescent microscope.

30. Cruz, J. C., Sugii, S., Yu, C. & Chang, T. Y. Role of Niemann-Pick type C1 protein in intracellular trafficking of low density lipoprotein-derived cholesterol. *J. Biol. Chem.* **275**, 4013–4021 (2000).
31. Geisbert, T. W. *et al.* Pathogenesis of Ebola hemorrhagic fever in cynomolgus macaques: evidence that dendritic cells are early and sustained targets of infection. *Am. J. Pathol.* **163**, 2347–2370 (2003).
32. Geisbert, T. W. *et al.* Pathogenesis of Ebola hemorrhagic fever in primate models: evidence that hemorrhage is not a direct effect of virus-induced cytolysis of endothelial cells. *Am. J. Pathol.* **163**, 2371–2382 (2003).
33. Whelan, S. P., Barr, J. N. & Wertz, G. W. Identification of a minimal size requirement for termination of vesicular stomatitis virus mRNA: implications for the mechanism of transcription. *J. Virol.* **74**, 8268–8276 (2000).
34. Whelan, S. P., Ball, L. A., Barr, J. N. & Wertz, G. T. Efficient recovery of infectious vesicular stomatitis virus entirely from cDNA clones. *Proc. Natl Acad. Sci. USA* **92**, 8388–8392 (1995).
35. Takada, A., Watanabe, S., Okazaki, K., Kida, H. & Kawaoka, Y. Infectivity-enhancing antibodies to Ebola virus glycoprotein. *J. Virol.* **75**, 2324–2330 (2001).
36. Carette, J. E. *et al.* Generation of iPSCs from cultured human malignant cells. *Blood* **115**, 4039–4042 (2010).
37. Morgenstern, J. P. & Land, H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.* **18**, 3587–3596 (1990).
38. Pentchev, P. G. *et al.* The cholesterol storage disorder of the mutant BALB/c mouse. A primary genetic lesion closely linked to defective esterification of exogenously derived cholesterol and its relationship to human type C Niemann-Pick disease. *J. Biol. Chem.* **261**, 2772–2777 (1986).
39. Ebert, D. H., Deussing, J., Peters, C. & Dermody, T. S. Cathepsin L and cathepsin B mediate reovirus disassembly in murine fibroblast cells. *J. Biol. Chem.* **277**, 24609–24617 (2002).
40. Blum, G. *et al.* Dynamic imaging of protease activity with fluorescently quenched activity-based probes. *Nature Chem. Biol.* **1**, 203–209 (2005).
41. Cureton, D. K., Massol, R. H., Saffarian, S., Kirchhausen, T. L. & Whelan, S. P. Vesicular stomatitis virus enters cells through vesicles incompletely coated with clathrin that depend upon actin for internalization. *PLoS Pathog.* **5**, e1000394 (2009).
42. Ehrlich, M. *et al.* Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell* **118**, 591–605 (2004).
43. Lefrançois, L. & Lyles, D. S. The interaction of antibody with the major surface glycoprotein of vesicular stomatitis virus. I. Analysis of neutralizing epitopes with monoclonal antibodies. *Virology* **121**, 157–167 (1982).
44. Wilson, J. A. *et al.* Epitopes involved in antibody-mediated protection from Ebola virus. *Science* **287**, 1664–1666 (2000).
45. Maruyama, T. *et al.* Ebola virus can be effectively neutralized by antibody produced in natural human infection. *J. Virol.* **73**, 6024–6030 (1999).
46. Bray, M., Davis, K., Geisbert, T., Schmaljohn, C. & Huggins, J. A mouse model for evaluation of prophylaxis and therapy of Ebola hemorrhagic fever. *J. Infect. Dis.* **178**, 651–661 (1998).
47. Warfield, K. L. *et al.* Development of a model for marburgvirus based on severe-combined immunodeficiency mice. *Virol. J.* **4**, 108 (2007).
48. Muhlberger, E., Weik, M., Volchkov, V. E., Klenk, H. D. & Becker, S. Comparison of the transcription and replication strategies of Marburg virus and Ebola virus by using artificial replication systems. *J. Virol.* **73**, 2333–2342 (1999).
49. Kranzusch, P. J. *et al.* Assembly of a functional Machupo virus polymerase complex. *Proc. Natl Acad. Sci. USA* **107**, 20069–20074 (2010).

Small molecule inhibitors reveal Niemann–Pick C1 is essential for Ebola virus infection

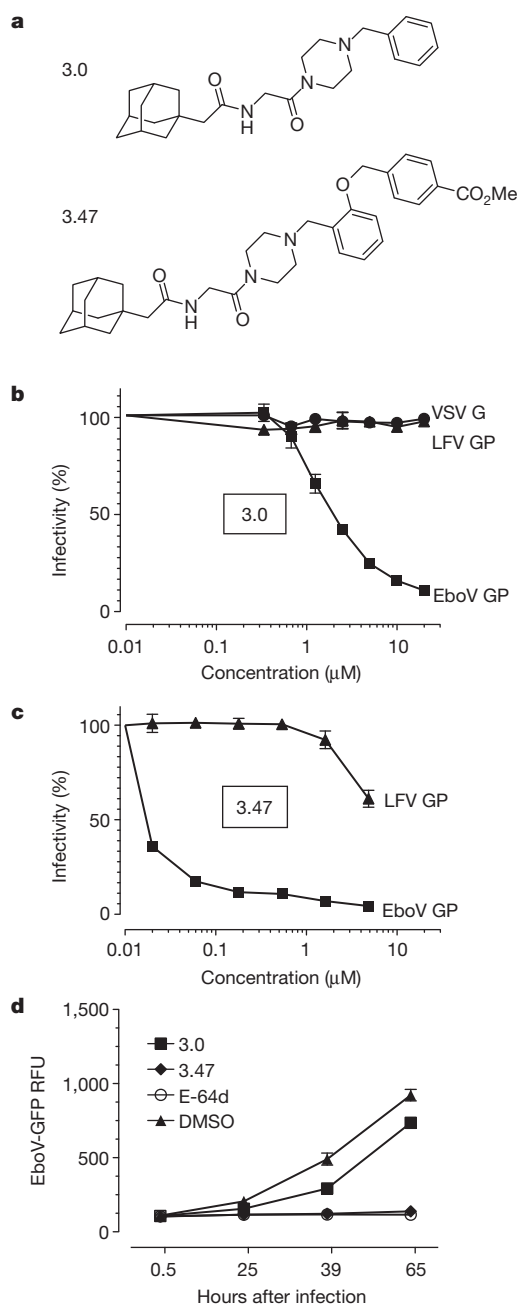
Marceline Côté^{1*}, John Misasi^{1,2*}, Tao Ren^{3*}, Anna Bruchez^{1*}, Kyungae Lee³, Claire Marie Filone^{1,4}, Lisa Hensley⁴, Qi Li¹, Daniel Ory⁵, Kartik Chandran^{1†} & James Cunningham^{1,6}

Ebola virus (EboV) is a highly pathogenic enveloped virus that causes outbreaks of zoonotic infection in Africa. The clinical symptoms are manifestations of the massive production of pro-inflammatory cytokines in response to infection¹ and in many outbreaks, mortality exceeds 75%. The unpredictable onset, ease of transmission, rapid progression of disease, high mortality and lack of effective vaccine or therapy have created a high level of public concern about EboV². Here we report the identification of a novel benzylpiperazine adamantane diamide-derived compound that inhibits EboV infection. Using mutant cell lines and informative derivatives of the lead compound, we show that the target of the inhibitor is the endosomal membrane protein Niemann–Pick C1 (NPC1). We find that NPC1 is essential for infection, that it binds to the virus glycoprotein (GP), and that antiviral compounds interfere with GP binding to NPC1. Combined with the results of previous studies of GP structure and function, our findings support a model of EboV infection in which cleavage of the GP1 subunit by endosomal cathepsin proteases removes heavily glycosylated domains to expose the amino-terminal domain^{3–7}, which is a ligand for NPC1 and regulates membrane fusion by the GP2 subunit⁸. Thus, NPC1 is essential for EboV entry and a target for antiviral therapy.

To identify chemical probes that target EboV host factors, we screened a library of small molecules and identified a novel benzylpiperazine adamantane diamide, 3.0, that inhibits infection of Vero cells by vesicular stomatitis virus particles (VSV) pseudotyped with EboV Zaire GP, but not with VSV G or Lassa fever virus (LFV) GP (Fig. 1a, b). To verify that 3.0 is a bona fide inhibitor, we measured EboV growth on Vero cells for 96 h and found it was reduced by >99% in the presence of 3.0 (Supplementary Fig. 1a). We synthesized and tested more than 50 analogues of 3.0 and found that the addition of a (methoxycarbonyl) benzyl group at the ortho position of the benzene ring (compound 3.47) increased the potency, as measured by a single cycle of EboV GP-dependent infection, and efficacy, as measured by growth of EboV on Vero cells (Fig. 1a, c, d).

Figure 1 | Structure and function of Ebola virus entry inhibitors.

a, Compounds 3.0 and 3.47. **b**, **c**, Vero cells were grown in media containing increasing concentrations of 3.0 (**b**) or 3.47 (**c**) for 90 min before the addition of VSV particles encoding luciferase (**b**) or GFP (**c**) and pseudotyped with either EboV GP, VSV G or Lassa fever virus GP (LFV GP). Virus infection is reported as percent of luminescence units (RLU) or GFP-positive cells relative to cells exposed to DMSO vehicle alone. Data are mean \pm s.d. ($n = 4$) and is representative of three experiments. **d**, Vero cells were grown in media containing 3.0 (40 μ M), 3.47 (40 μ M), vehicle (1% DMSO) or the cysteine cathepsin protease inhibitor E-64d (150 μ M) 90 min before the addition of replication competent Ebola virus Zaire-Mayinga encoding GFP (multiplicity of infection (m.o.i.) = 0.1). Results are mean relative fluorescence units (RFU) \pm s.e.m. ($n = 3$).



¹Division of Hematology, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ²Division of Infectious Disease, Department of Medicine, Children's Hospital, Boston, Massachusetts 02115, USA. ³New England Regional Center of Excellence for Biodefense and Emerging Infectious Diseases, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴United States Army Medical Research Institute of Infectious Diseases, Virology Division, Frederick, Maryland 21702, USA. ⁵Diabetic Cardiovascular Disease Center, Washington University School of Medicine, Saint Louis, Missouri 63110, USA. ⁶Department of Microbiology and Immunology, Harvard Medical School, Boston, Massachusetts 02115, USA. †Present address: Department of Microbiology and Immunobiology, Albert Einstein College of Medicine, Bronx, New York 10461, USA.

*These authors contributed equally to this work.

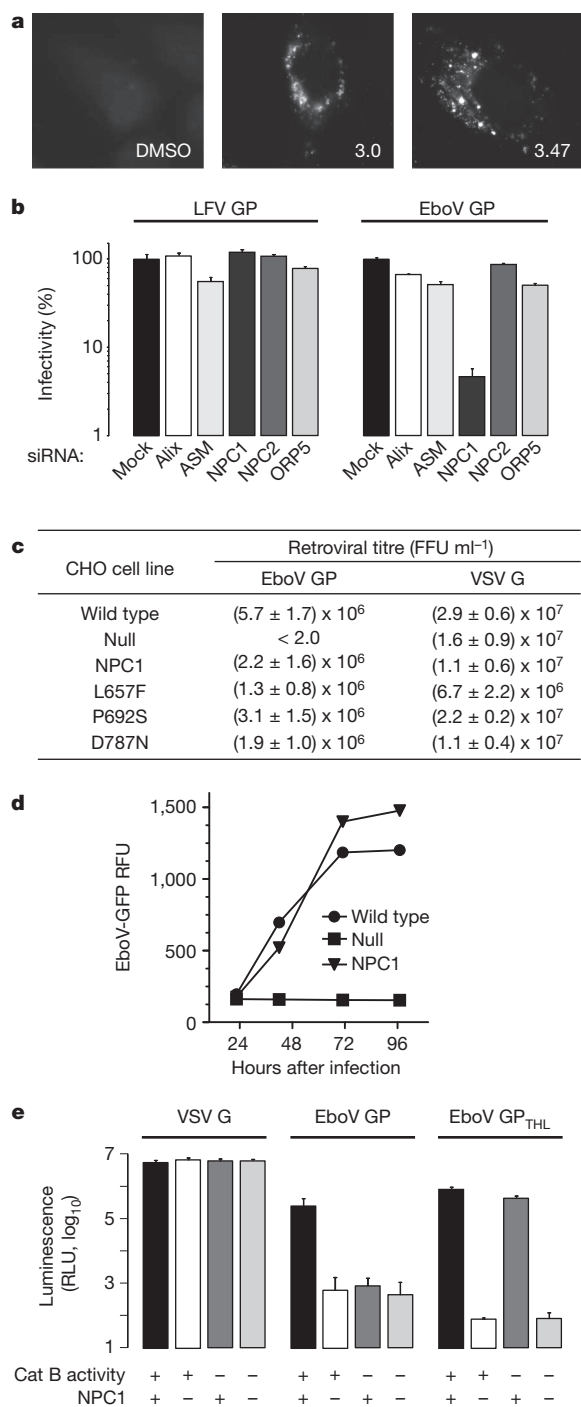


Figure 2 | NPC1 is essential for Ebola virus infection. **a**, HeLa cells were treated with 3.0 (20 μ M), 3.47 (1.25 μ M) or vehicle for 18 h, then fixed and incubated with the cholesterol-avid fluorophore filipin. **b**, HeLa cells were transfected with siRNAs targeting Alix, ASM, NPC1, NPC2 and ORP5. After 72 h, VSV EboV GP or LFV GP infection of these cells was measured as in Fig. 1c. Data are mean \pm s.d. ($n = 3$) and is representative of three experiments. **c**, CHO_{wt}, CHO_{null} and CHO_{NPC1} cells stably expressing mouse NPC1 (CHO_{NPC1}) or NPC1 mutants L657F, P692S, D787N were exposed to MLV particles encoding LacZ and pseudotyped with either EboV GP or VSV G. Results are the mean \pm s.d. ($n = 4$) and is representative of three experiments. FFU, focus forming units. **d**, CHO_{wt}, CHO_{null} and CHO_{NPC1} cells were infected with replication competent Ebola virus Zaire-Mayinga encoding GFP (m.o.i. = 1). Results are mean relative fluorescence units \pm s.d. ($n = 3$). **e**, CHO_{wt} and CHO_{null} cells were treated with the cathepsin B inhibitor CA074 (80 μ M) or vehicle. These cells were challenged with VSV G particles or VSV EboV GP particles treated with thermolysin (EboV GP_{THL}) or untreated control (EboV GP). Infection was measured as in Fig. 1b. Data are mean \pm s.d. ($n = 9$).

Previous studies revealed that the endosomal protease cathepsin B is essential for EboV infection because it cleaves the GP1 subunit of GP^{3,4}. To address the possibility that 3.0 and 3.47 target this step, we measured cathepsin B activity in the presence of these compounds and found no effect *in vitro* or in cells (data not shown). Moreover, 3.0 and 3.47 inhibited infection by VSV EboV particles treated with thermolysin, a metalloprotease that faithfully mimics cathepsin cleavage of the GP1 subunit of GP (Supplementary Fig. 1b)^{4,9}. These findings demonstrate that cathepsin B is not the target of 3.0 and 3.47.

HeLa cells treated with 3.0 or 3.47 for more than 18 h developed cytoplasmic vacuoles that were labelled by cholesterol-avid filipin (Fig. 2a). The induction of filipin-stained vacuoles by the compounds

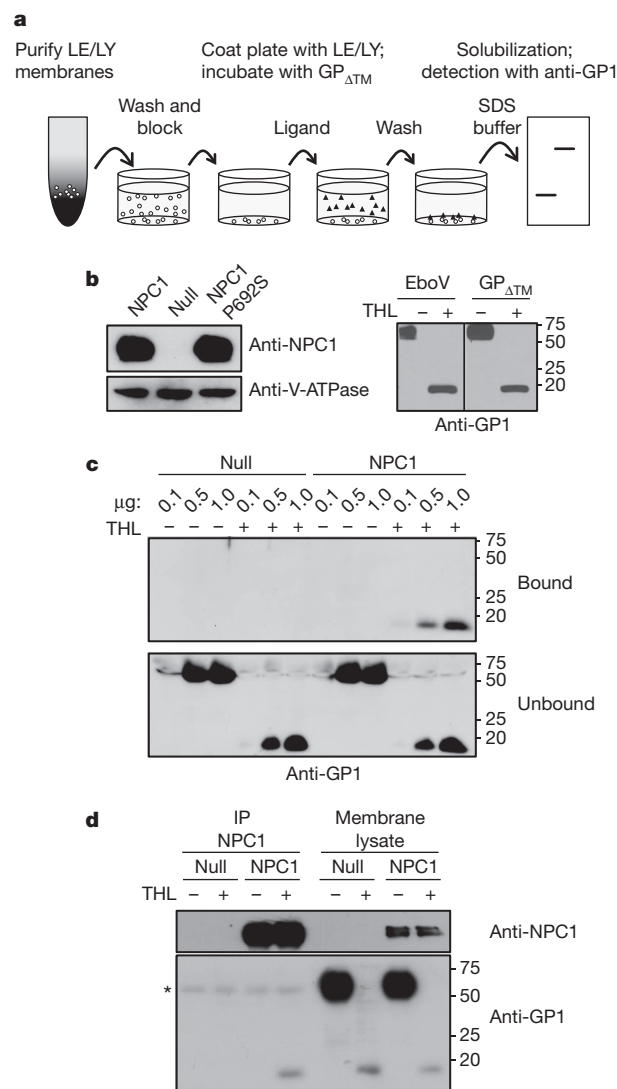


Figure 3 | Protease-cleaved EboV GP binds to NPC1. **a**, Schematic diagram of EboV GP1 binding assay used in panel c. **b**, left, LE/LY membranes from CHO_{NPC1}, CHO_{null} and CHO NPC1 P692S cells were analysed by immunoblot using antibodies to NPC1 or V-ATPase B1/2. Right, VSV-EboV GP particles and EboV GP_{ΔTM} protein were incubated in the presence or absence of thermolysin (THL) and analysed by immunoblot for GP1. **c**, EboV GP_{ΔTM} or thermolysin-cleaved EboV GP_{ΔTM} (0.1, 0.5, or 1.0 μ g) was added to LE/LY membranes purified from CHO_{null} or CHO_{NPC1} cells. Membrane bound and unbound GP1 were analysed by immunoblot. **d**, LE/LY membranes from CHO_{null} or CHO_{NPC1} cells were incubated with EboV GP_{ΔTM} or thermolysin-cleaved EboV GP_{ΔTM}. Following binding, membranes were dissolved in CHAPSO, NPC1 was precipitated using an NPC1-specific antibody, and the immunoprecipitate and the input membrane lysate were analysed by immunoblot for NPC1 (top) or GP1 (bottom). * IgG heavy chain.

indicated that they target one or more proteins involved in regulation of cholesterol uptake in cells. To test this hypothesis, we used mutant cell lines and cells treated with small interfering RNA (siRNA) to analyse proteins for which loss of activity had been previously associated with cholesterol accumulation in late endosomes^{10–12}. We found that EboV GP infection is dependent on the expression of Niemann–Pick C1 (NPC1), but not Niemann–Pick C2 (NPC2), acid sphingomyelinase (ASM), ALG-2-interacting protein X (Alix), or oxysterol binding protein 5 (ORP5) (Fig. 2b, Supplementary Fig. 2a–c). NPC1 is a polytopic protein that resides in the limiting membrane of late endosomes and lysosomes (LE/LY) and mediates distribution of lipoprotein-derived cholesterol in cells^{10,13}. To analyse the role of NPC1 in infection, we studied Chinese hamster ovary (CHO)-derived cell lines that differ in expression of NPC1. We found that the titre of a murine leukaemia virus (MLV) vector pseudotyped with EboV GP on wild-type CHO cells (CHO_{wt}) exceeded 10⁶ infectious units per ml (Fig. 2c). Importantly, CHO cells lacking NPC1 (CHO_{null}) were completely resistant to infection by this virus and infection of these cells was fully restored when NPC1 was expressed (CHO_{NPC1}). Thus, NPC1 expression is essential for EboV infection.

In CHO_{null} cells, LE/LY are enlarged and contain excess cholesterol (Supplementary Fig. 3)¹⁴. To determine if EboV infection is inhibited by endosome dysfunction secondary to the absence of NPC1, we studied a well-characterized NPC1 mutant P692S that is defective in cholesterol uptake and NPC1-dependent membrane trafficking^{13–15} and found that expression of NPC1 P692S fully supports infection of CHO_{null} cells (Fig. 2c). Conversely, gain-of-function mutants NPC1 L657F and NPC1 D787N (ref. 14) did not enhance EboV GP infection. Thus, EboV entry is strictly dependent on NPC1 expression, but not NPC1-dependent cholesterol transport activity. Consistent with the conclusion that NPC1 expression is essential for EboV

GP-dependent entry, we found that Ebola virus did not grow on CHO_{null} cells (Fig. 2d). In addition, we tested a single round of infection by MLV particles bearing GPs from the filoviruses EboV Sudan, EboV Côte d'Ivoire, EboV Bundibugyo, EboV Reston and Marburg virus and found that all are strictly NPC1-dependent (Supplementary Fig. 4). Because these viruses are not closely related¹⁶, these findings indicate that the requirement for NPC1 as an entry factor is conserved among viruses in the *Filoviridae* family.

Because NPC1 and cathepsin B are both essential host factors, we analysed their relationship during infection. In our initial experiment, we measured cathepsin B activity in CHO_{null} cells and found it was not significantly different from CHO_{wt} cells (Supplementary Fig. 5). To determine if NPC1 is required for virus processing by cathepsin B, we tested whether thermolysin-cleaved particles are dependent on NPC1. As expected, we found that thermolysin-cleaved particles are infectious and resistant to inactivation of cathepsin B when NPC1 is present (Fig. 2e). However, thermolysin cleavage did not bypass the barrier to virus infection of NPC1 deficient cells. Taken together, these findings indicate that cathepsin B and NPC1 mediate distinct steps in infection.

Previous studies suggest that the product of cathepsin B cleavage of the GP1 subunit of EboV GP is a ligand for a host factor^{6,17–20}. To test this hypothesis, we performed a series of experiments measuring binding of EboV GP to LE/LY membranes from CHO_{null}, CHO_{NPC1} and CHO_{P692S} cells (Fig. 3a, b, left panel). The source of EboV GP is a purified recombinant protein that is truncated just before the trans-membrane domain (EboV GP_{ΔTM}). EboV GP_{ΔTM} is a trimer that is faithfully cleaved by thermolysin (Fig. 3b, right panel). We found that binding of EboV GP_{ΔTM} to LE/LY membranes is concentration-dependent, saturable, and strictly dependent on both thermolysin cleavage of GP1 and membrane expression of NPC1 or NPC1 P692S (Fig. 3c and Supplementary Fig. 6a, b). To determine if cleaved GP

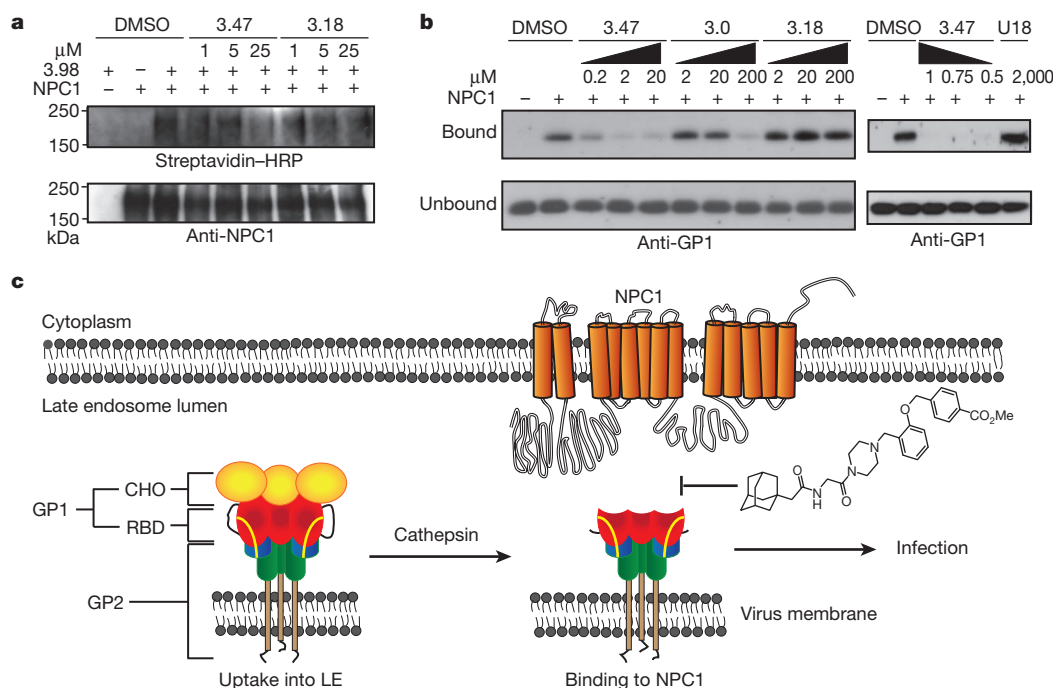


Figure 4 | NPC1 is a target of the small molecule inhibitors. **a**, LE/LY membranes from CHO_{null} or CHO_{NPC1} cells were incubated at the indicated concentrations of 3.47, 3.18 or DMSO (5%) before the addition of the photoactivatable 3.98 (25 μM). After incubation, 3.98 was activated by ultraviolet light and then conjugated to biotin. NPC1 was immunoprecipitated and analysed by immunoblot for conjugation of 3.98 to NPC1 using streptavidin–horseradish peroxidase (HRP) (top) and recovery of NPC1 (bottom). **b**, Thermolysin-cleaved EboV GP_{ΔTM} protein (1 μg) was added to LE/LY membranes from CHO_{null} or CHO_{NPC1} cells in the presence of DMSO (10%) or

the indicated concentrations of 3.47, 3.0, or 3.18 (left panel), and 3.47 or U18666A (U18, right panel). Membrane-bound and unbound GP1 were analysed by immunoblot. **c**, Proposed model of EboV entry. Following EboV uptake and trafficking to late endosomes^{24,25}, EboV GP is cleaved by cathepsin protease to remove heavily glycosylated domains (CHO) and expose the putative receptor binding domain (RBD) of GP1 (refs 6, 17–19). Binding of cleaved GP1 to NPC1 is necessary for infection and is blocked by the EboV inhibitor 3.47.

binds to NPC1, we performed a co-immunoprecipitation experiment. LE/LY membranes were incubated with EboV GP_{ATM} and then solubilized in detergent. NPC1 was recovered from the lysate by immunoprecipitation and the immune complexes were analysed for GP1. The findings indicate that cleaved EboV GP_{ATM} binds to NPC1 and that uncleaved EboV GP_{ATM} does not (Fig. 3d).

Because the small molecules 3.0 and 3.47 inhibit infection of thermolysin-treated VSV EboV GP particles (Supplementary Fig. 1b) and inhibit cholesterol uptake from LE/LY into cells (Fig. 2a), both of which require NPC1, this suggests the possibility that these compounds directly target NPC1. To test this hypothesis, we synthesized the 3.47 derivative 3.98. This compound has anti-EboV activity and contains two additional functional moieties: an aryl-azide for photoaffinity labelling of target proteins and an alkyne for click conjugation with biotin²¹ (Supplementary Fig. 7). Compound 3.98 was incubated with LE/LY membranes, activated by ultraviolet light and coupled to biotin. NPC1 was then isolated by immunoprecipitation and analysed using streptavidin–horseradish peroxidase. The findings show that NPC1 is cross-linked to 3.98 and that cross-linking is inhibited by the presence of 3.47 but not by the closely related analogue 3.18, which has weak antiviral activity (Fig. 4a and Supplementary Fig. 7). In addition, we observed that overexpression of NPC1 conferred resistance to the antiviral activity of 3.0 and 3.47 (Supplementary Fig. 8), thus providing additional functional evidence supporting the conclusion based on the results of the cross-linking experiment using 3.98 that NPC1 is a direct target of the antiviral compounds.

The evidence that NPC1 is the target of the 3.0-derived small molecules selected for anti-EboV activity indicated that these compounds may interfere with binding of cleaved GP to NPC1. Consistent with this hypothesis, we found that 3.0 and 3.47 inhibited binding of cleaved EboV GP_{ATM} to NPC1 membranes in a concentration-dependent manner (Fig. 4b). Importantly, we observed a direct correlation between the potency of 3.47, 3.0 and 3.18 in inhibiting binding (Fig. 4b, left panel) and in inhibiting EboV infection (Supplementary Fig. 7). We also tested U18666A, a small molecule inhibitor of LE/LY cholesterol transport and membrane trafficking^{22,23}, and found that it does not inhibit binding of cleaved EboV GP to NPC1 membranes (Fig. 4b, right panel). These results support the conclusion that the 3.0-derived compounds inhibit EboV infection by interfering with binding of cleaved GP to NPC1.

Previous studies show that cleavage of GP by endosomal cathepsin proteases removes heavily glycosylated domains in the GP1 subunit and exposes the N-terminal domain^{3–7}. It has been proposed that binding of this domain to a host factor is essential for infection^{6,17–20}. The most straightforward interpretation of the findings in this report is that NPC1 is this host factor. This conclusion is based on the observations that NPC1 is strictly required for infection, that cleaved GP1 binds to NPC1, and that small molecules that target NPC1 are potent inhibitors of binding and infection.

Analysis of the EboV GP structure shows that the residues in the N-terminal domain of GP1 that mediate binding to NPC1 are interspersed with the residues that make stabilizing contacts with GP2 (ref. 5). This structural feature is consistent with the possibility that binding of cleaved GP1 to NPC1 relieves the GP1-imposed constraints on GP2 and promotes virus fusion to the limiting membrane (Fig. 4c). The role of cathepsin proteases in cleavage of GP1 to expose the NPC1 binding site during EboV infection is analogous to the role of CD4 in inducing a conformational change in gp120 to expose the co-receptor binding site during human immunodeficiency virus infection⁸. An alternative possibility is that binding of protease-cleaved GP1 to NPC1 is an essential step in infection, but virus membrane fusion is not completed until an additional signal is received, possibly including further cleavage of GP by cathepsin proteases, as has been proposed^{3,4,9}. These studies provide an example of how small molecules identified by screening and medicinal chemistry optimization can be used as molecular probes to analyse virus–host interactions.

METHODS SUMMARY

Screening of small molecules was performed at the New England Regional Centers of Excellence for Biodefense and Emerging Infectious Diseases at Harvard Medical School. Infection was assayed using VSV pseudotyped viruses encoding green fluorescent protein (GFP) or luciferase. Experiments with native Ebola virus were performed under BSL-4 conditions at the United States Army Medical Research Institute for Infectious Diseases. Cells were infected with EboV Zaire-Mayinga GFP and growth was measured by mean fluorescence. EboV GP_{ATM} is a derivative of EboV GP in which the transmembrane domain has been replaced by a GCN4-derived trimerization domain followed by a His₆ tag for purification. Late endosomes/lysosomes (LE/LY) were isolated by differential centrifugation and further purified by Percoll density gradient centrifugation. LE/LY were disrupted by incubation with methionine methyl ester and coated onto high binding ELISA plates. Following attachment, unbound LE/LY membranes were removed and plates were blocked. Bound membranes were incubated with the indicated amounts of native or thermolysin-cleaved EboV GP_{ATM} protein. Unbound EboV GP_{ATM} protein was removed, membranes were washed and bound EboV GP_{ATM} protein was recovered in SDS loading buffer and analysed by immunoblot using GP1 antiserum. Where applicable, membranes were pre-incubated with 3.0, 3.47, 3.18 or vehicle before the addition of EboV GP_{ATM}. To analyse EboV GP_{ATM} binding to NPC1, LE/LY membranes were dissolved in 10 mM CHAPS, NPC1 was recovered by immunoprecipitation, and the immune complexes were analysed by immunoblot using GP1 antiserum.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 March; accepted 19 July 2011.

Published online 24 August 2011.

1. Zampieri, C. A., Sullivan, N. J. & Nabel, G. J. Immunopathology of highly virulent pathogens: insights from Ebola virus. *Nature Immunol.* **8**, 1159–1164 (2007).
2. Geisbert, T. W. & Jahrling, P. B. Exotic emerging viral diseases: progress and challenges. *Nature Med.* **10**, S110–S121 (2004).
3. Chandran, K., Sullivan, N. J., Felber, U., Whelan, S. P. & Cunningham, J. M. Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection. *Science* **308**, 1643–1645 (2005).
4. Schornberg, K. *et al.* Role of endosomal cathepsins in entry mediated by the Ebola virus glycoprotein. *J. Virol.* **80**, 4174–4178 (2006).
5. Lee, J. E. *et al.* Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature* **454**, 177–182 (2008).
6. Dube, D. *et al.* The primed ebolavirus glycoprotein (19-kilodalton GP_{1,2}): Sequence and residues critical for host cell binding. *J. Virol.* **83**, 2883–2891 (2009).
7. Hood, C. L. *et al.* Biochemical and structural characterization of cathepsin L-processed Ebola virus glycoprotein: Implications for viral entry and immunogenicity. *J. Virol.* **84**, 2972–2982 (2010).
8. Harrison, S. C. Viral membrane fusion. *Nature Struct. Mol. Biol.* **15**, 690–698 (2008).
9. Wong, A., Sandesara, R., Mulherkar, N., Whelan, S. & Chandran, K. A forward genetic strategy reveals destabilizing mutations in the ebolavirus glycoprotein that alter its protease dependence during cell entry. *J. Virol.* **84**, 163–175 (2010).
10. Kolter, T. & Sandhoff, K. Lysosomal degradation of membrane lipids. *FEBS Lett.* **584**, 1700–1712 (2010).
11. Du, X. *et al.* A role for oxysterol-binding protein-related protein 5 in endosomal cholesterol trafficking. *J. Cell Biol.* **192**, 121–135 (2011).
12. Chevallier, J. *et al.* Lyso-bisphosphatidic acid controls endosomal cholesterol levels. *J. Biol. Chem.* **283**, 27871–27880 (2008).
13. Ko, D. C., Gordon, M. D., Jin, J. Y. & Scott, M. P. Dynamic movements of organelles containing Niemann-Pick C1 protein: NPC1 involvement in late endocytic events. *Mol. Biol. Cell* **12**, 601–614 (2001).
14. Millard, E. E. *et al.* The sterol-sensing domain of the Niemann-Pick C1 (NPC1) protein regulates trafficking of low density lipoprotein cholesterol. *J. Biol. Chem.* **280**, 28581–28590 (2005).
15. Ohgami, N. *et al.* Binding between the Niemann-Pick C1 protein and a photoactivatable cholesterol analog requires a functional sterol-sensing domain. *Proc. Natl Acad. Sci. USA* **101**, 12473–12478 (2004).
16. Towner, J. S. *et al.* Newly discovered Ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLoS Pathog.* **4**, e1000212 (2008).
17. Kuhn, J. H. *et al.* Conserved receptor-binding domains of Lake Victoria marburgvirus and Zaire ebolavirus bind a common receptor. *J. Biol. Chem.* **281**, 15951–15958 (2006).
18. Kaletsky, R. L., Simmons, G. & Bates, P. Proteolysis of the Ebola virus glycoproteins enhances virus binding and infectivity. *J. Virol.* **81**, 13378–13384 (2007).
19. Brindley, M. A. *et al.* Ebola virus glycoprotein 1: Identification of residues important for binding and postbinding events. *J. Virol.* **81**, 7702–7709 (2007).
20. Dube, D. *et al.* Cell adhesion-dependent membrane trafficking of a binding partner for the ebolavirus glycoprotein is a determinant of viral entry. *Proc. Natl Acad. Sci. USA* **107**, 16637–16642 (2010).
21. Ban, H. S. *et al.* Identification of HSP60 as a primary target of o-carboranylphenylphenoxyacetanilide, an HIF-1 α inhibitor. *J. Am. Chem. Soc.* **132**, 11870–11871 (2010).

22. Sobo, K. *et al.* Late endosomal cholesterol accumulation leads to impaired intra-endosomal trafficking. *PLoS ONE* **2**, e851 (2007).
23. Huynh, K. K., Gershenson, E. & Grinstein, S. Cholesterol accumulation by macrophages impairs phagosome maturation. *J. Biol. Chem.* **283**, 35745–35755 (2008).
24. Saeed, M. F., Kolokoltsov, A. A., Albrecht, T. & Davey, R. A. Cellular entry of Ebola virus involves uptake by a macropinocytosis-like mechanism and subsequent trafficking through early and late endosomes. *PLoS Pathog.* **6**, e1001110 (2010).
25. Nanbo, A. *et al.* Ebolavirus is internalized into host cells via macropinocytosis in a viral glycoprotein-dependent manner. *PLoS Pathog.* **6**, e1001121 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Considine, A. Nilsson and S. Wilkes for assistance, S. Chiang for critical reading of the manuscript, G. Beltz, N. Gray, S. Grinstein, Y. Iannou, R. Infante, J. Kornhuber, F. Sharom and S. Whelan for discussion. This work was supported by grants from U54 AI057159, R01 CA104266 to J.C., PIDS-Sanofi-Pasteur Fellowship, K12-HD052896 and 5K08AI079381 to J.M., 5-T32-HL007623 to A.B., and

fellowship from Fonds de la Recherche en Santé du Québec to M.C.; C.M.F. was supported by the Postgraduate Research Participation Program at the US Army Medical Research and Materiel Command administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and USAMRMC.

Author Contributions M.C., J.M., T.R. and A.B. equally contributed to this work. K.C. and T.R. performed the inhibitor screen. K.L. synthesized and purified 3.0 analogues and T.R. tested them. T.R., A.B., J.M., Q.L. and M.C. carried out infection assays with pseudotyped viruses. A.B. performed microscopy. J.M. purified recombinant glycoprotein. M.C. and J.M. designed and performed binding assays. M.C. performed immunoprecipitation. D.O. provided NPC1 constructs, antibodies and CHO cell lines. Ebola virus infections were performed in the lab of L.H. by C.M.F.; J.C. supervised the project and wrote the manuscript. All authors reviewed the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.C. (jcunningham@rics.bwh.harvard.edu).

METHODS

Cell lines. Vero, 293T, HeLa (ATCC) and human fibroblasts²⁶ (Coriell) were maintained in DMEM (Invitrogen) supplemented with 5% FetalPlex, 5% FBS (Gemini) or 10% FBS (HeLa, human fibroblasts). All CHO derived cell lines were grown as previously described^{14,27}. We have designated the CHO-K1 cell line as CHO_{wt}, CHO-M12 as CHO_{null}, CHO-wt8 as CHO_{NPC1}, and the CHO-derived cell lines expressing NPC1 mutants as CHO NPC1 P692S, CHO NPC1 L657F, and CHO NPC1 D787N. CHO/NPC1-1, designated here as CHO hNPC1, expresses high levels of human NPC1 (ref. 27).

Antibodies. Rabbit polyclonal anti-serum was raised against a peptide corresponding to residues 83 to 98 of Ebola virus Zaire Mayinga GP1 (TKRWGFRSGVPPKVVC). Antibodies to NPC1 and V-ATPase B1/2 were obtained from Abcam and Santa Cruz, respectively.

Expression plasmids. Mucin domain-deleted EboV Zaire Mayinga GP (EboV GP) and VSV G were previously described³. Plasmids encoding Côte d'Ivoire-Ivory Coast GP, Sudan-Boniface GP, Reston-Penn. GP and Marburg-Musoke GP were obtained from Anthony Sanchez and the mucin domain-deleted (Δ Muc) derivatives were created: Zaire Δ Muc GP (amino acids 309–489 deleted), Côte d'Ivoire Δ Muc GP (amino acids 310–489 deleted), Sudan Δ Muc GP (Δ a.a. 309–490), and Reston Δ Muc GP amino acids 310–490 deleted). Bundibungyo-Uganda viral RNA was TRIzol-extracted and PCR was used to generate a construct that expresses a mucin-deleted GP (amino acids 309–489 deleted). A plasmid encoding Lassa fever virus GP1 was kindly provided by G. Nabel. A codon-optimized sequence encoding GP2 was generated and combined with the GP1 sequence in pCAGGS to complete a GP expression vector.

Production and purification of pseudotyped virions. VSV- Δ G pseudotyped viruses were created as described previously³. LacZ-encoding retroviral pseudotypes bearing the designated envelope glycoproteins were prepared as previously described²⁸.

Thermolysin digestion of EboV GP Virus and EboV GP_{ATM}. Purified EboV GP_{ATM} (50 μ g ml⁻¹) or VSV particles pseudotyped with EboV GP were incubated at 37 °C for 1 h with the metalloprotease thermolysin (Sigma, 0.2 mg ml⁻¹) in NT buffer (10 mM Tris-HCl pH 7.5, 135 mM NaCl). The reaction was stopped using 500 mM phosphoramidon (Sigma) at 4 °C. Cleaved EboV GP_{ATM} was stored in phosphate buffered saline supplemented with 1 mM EDTA, 1 mM PMSF (Sigma) and 1× EDTA-Free Complete Protease Inhibitor Cocktail (Roche).

Infection assays with pseudotyped virus. VSV pseudotyped viruses expressing GFP were added to cells in serial tenfold dilutions and assayed using fluorescence microscopy. An infectious unit (i.u.) is defined as one GFP-expressing cell within a range where the change in GFP-positive cells is directly proportional to the virus dilution. For VSV expressing the luciferase reporter, pseudotyped virus was added to cells and luciferase activity was assayed 6–20 h post-infection using the firefly luciferase kit (Promega). Signal was measured in relative luminescence units (RLU) using an EnVision plate reader (Perkin Elmer). In experiments involving inhibitors, stock solutions of 3.0 (20 mM) and 3.47 (10 mM) in DMSO were diluted to a final concentration of 1% DMSO in media. Inhibitory activity was stable in the media of cultured cells for more than 72 h as assessed using a single cycle entry assay. Infection of target cells with LacZ-encoding retroviral pseudotypes was performed in the presence of 5 μ g ml⁻¹ polybrene (Sigma). Seventy-two hours post-infection, cells were stained for LacZ activity and titre was determined by counting positive foci and expressed as focus forming units (FFU) per ml of virus.

Ebola virus infections under BSL-4 conditions. Vero cells or CHO cells were seeded to 96-well plates and exposed to EboV-GFP²⁹. Vero cells were incubated with 3.0 (40 μ M), 3.47 (40 μ M), E-64-d (150 μ M) or 1% DMSO 90 min before the addition of virus (m.o.i. = 0.1). Virus was added to CHO cells at m.o.i. of 1 as measured on Vero cells. Virus-encoded GFP fluorescence was determined using a SpectraMax M5 plate reader (Molecular Devices) at excitation 485 nm, emission 515 nm, cutoff 495 nm at 22.5, 42, 71 and 97 h post-infection. An additional inhibitor experiment was performed using 3.0. Vero cells were treated with 3.0 (20 μ M) or 1% DMSO alone for 4 h, and then infected with EBOV Zaire-1995 (m.o.i. = 0.1). After 1 h, the virus inoculum was removed, cells were washed, and fresh media containing 3.0 or DMSO was added. Cell supernatant was collected at 0, 24, 48, 72, or 91 h post-infection. RNA was isolated from the supernatant using Virus RNA Extraction kits (Qiagen) and EboV NP RNA was measured using a real-time RT-PCR assay³⁰. Virus titre was calculated using a standard curve obtained using a virus stock of known titre as determined by plaque assay.

Screen for Ebola virus entry inhibitors. Screening of small molecules was performed at the New England Regional Centers of Excellence for Biodefense and Emerging Infectious Diseases at Harvard Medical School. Vero cells were seeded in 384-well plates at a density of 5×10^3 cells per well using a Matrix WellMate (Thermo Scientific). The ChemBridge3, ChemDiv4, ChemDiv5 and Enamine2 compound libraries were transferred by robotics to the assay plates using stainless

steel pin arrays. The compounds were screened at a constant dilution to achieve a final concentration between 10 μ M and 60 μ M. After incubation for 2 h at 37 °C, viruses were dispensed into each well (m.o.i. = 1) and incubated for an additional 6 h to allow virus gene expression. Cells were lysed by addition of Steady-Glo (Promega) and after 10 min at room temperature luminescence was measured using an EnVision plate reader. Each compound was tested in duplicate. Candidate compounds that inhibited EboV GP infection by more than 80% were analysed for potency, selectivity and absence of cytotoxicity (using Cyto-Tox assay, Promega) and 3.0 (2-((3r,5r,7r)-adamantan-1-yl)-N-(2-(4-benzylpiperazin-1-yl)-2-oxoethyl)acetamide) was identified. The antiviral activity of the inhibitors was verified on human cells (HeLa, A549, 293T), mouse embryonic fibroblasts and Chinese hamster ovary cells.

Synthesis of 3.0 derivatives. Compound 3.47 (methyl 4-((2-((4-(2-((3r,5r,7r)-adamantan-1-yl)acetamido)acetyl)piperazin-1-yl)methyl)phenoxy)methyl)benzoate) was prepared via a multi-step synthesis starting from N-Cbz-piperazine. Thus, coupling of N-Cbz-piperazine with N-Boc-glycine followed by removal of the Boc group under acidic conditions yielded 4-Cbz-piperazine glycineamide. After acylation of the terminal amine with adamantan-1-acetyl chloride, the Cbz group was removed by hydrogenolysis to give (1-(adamantan-1-yl)acetamido)acetyl piperazine. The piperazine was then benzylated via reductive amination with 2-(4-methoxycarbonyl)benzyloxybenzaldehyde using sodium triacetoxyborohydride to provide 3.47. Compound 3.18 was synthesized in a similar fashion. Compound 3.98 was prepared via a multi-step synthesis as follows. First, 2-hydroxy-5-nitrobenzaldehyde was alkylated by 4-ethynylbenzyl bromide in the presence of potassium carbonate in DMF. Resulting benzyloxy aldehyde underwent reductive amination with 2-((3r,5r,7r)-adamantan-1-yl)-N-(2-oxo-2-(piperazin-1-yl)ethyl)acetamide using sodium triacetoxyborohydride. The nitro group was then reduced to aniline (SnCl₂), diazotized (NaNO₂), and the diazonium finally converted to azide to yield 3.98. See Supplementary Information for detailed experimental procedures and characterization data.

Protease inhibitors and protease activity assays. The measurement of cathepsin B activity and the use of the inhibitor CA074 (Sigma) have been previously described³.

Detection of intracellular cholesterol. Cells were stained with filipin (50 μ g ml⁻¹, Cayman Chemical) as previously described¹⁴. Images of stained cells were obtained using epifluorescence microscopy (Nikon Eclipse TE2000U). The images in the supplementary figures were processed using ImageJ software.

Production and purification of EboV GP_{ATM} soluble protein. EboV GP_{ATM} is a derivative of the mucin-deleted EboV Zaire-Mayinga GP in which the trans-membrane domain and carboxy terminus (amino acids 657–676) has been replaced by a GCN4-derived trimerization domain (MKQIEDKIEILSKIYHIEN EIARIKKLIGEV) and a His₆ tag. The expression plasmid encoding EboV GP_{ATM} was transfected into 293T cells using lipofectamine2000. Eighteen to twenty-four hours later the culture medium was replaced with 293FSFMI (Invitrogen) supplemented with 1× non-essential amino acids and 2 mM CaCl₂ and collected daily for 4 days. Media containing soluble EboV GP_{ATM} was filtered and PMSF (1 mM)/1× EDTA-Free Complete Protease Inhibitor Cocktail was added. EboV GP_{ATM} was purified by affinity chromatography using Ni-NTA agarose beads (Qiagen), dialysed against PBS using a 3 kDa dialysis cartridge (Pierce) and stored at –80 °C. Purity and integrity of EboV GP_{ATM} were analysed by SDS-PAGE.

Membrane binding assay. Indicated cells were washed with PBS twice, scraped in homogenization (HM) buffer (0.25 M sucrose, 1 mM EDTA, 10 mM HEPES pH 7.0), and disrupted with a Dounce homogenizer. Nuclei and debris were pelleted by centrifugation at 1,000g for 10 min. The post-nuclear supernatant was centrifuged at 15,000g for 30 min at 4 °C and the pellet, containing the LE/LY, was resuspended in a total volume of 0.9 ml composed of 20% Percoll (Sigma) and 0.4% BSA (Sigma) in HM and centrifuged at 36,000g for 30 min at 4 °C. Fractions (0.150 ml) were collected from the bottom to the top of the tube and those containing the highest β -N-acetylglucosamidase activity, as assessed by release of 4-methylumbelliferone from 4-methylumbelliferyl-N-acetyl- β -D-glucosaminide (Sigma), were pooled and incubated in 20 mM methionine methyl-ester (Sigma) for 1 h at room temperature. Following LE/LY disruption, 1× EDTA-Free Complete Protease Inhibitor Cocktail and 1 mM PMSF was added. The amount of purified LE/LY membranes used for the binding assay was normalized using the activity of the marker β -N-acetylglucosamidase and validated by immunoblot using V-ATPase B1/2 antibody (Supplementary Fig. 5).

Disrupted LE/LY membranes were coated on high-binding ELISA plates (Corning) overnight at 4 °C. Unbound membranes were removed and wells containing bound membranes were blocked for 2 h at room temperature with binding buffer (PBS, 5% FBS, 1 mM PMSF, 1 mM EDTA, 1× Complete Protease Inhibitor Cocktail). The indicated amount of purified EboV GP_{ATM}, pretreated or not with thermolysin, in binding buffer was added to each well and incubated for 1 h at room temperature. Unbound proteins were removed and wells were washed three

times with PBS. Membrane-bound EboV GP_{ATM} was solubilized in SDS-loading buffer. Bound and unbound EboV GP_{ATM} were detected by immunoblot using the EboV GP1 anti-serum. For binding assays in the presence of inhibitors, the immobilized membranes were pre-incubated at room temperature with the inhibitor or vehicle (10% DMSO) in binding buffer. After 30 min, thermolysin-cleaved EboV GP_{ATM} was added in the continuous presence of compound and bound and unbound GP was measured as described above.

Co-immunoprecipitation. CHO_{null} and CHO_{hNPC1} cells were homogenized as described above. The 15,000g membrane pellet was resuspended in HM buffer and protein content was measured using the BCA assay (Pierce). The LE/LY membranes contained in the 15,000g resuspended pellet were disrupted by incubation with 20 mM methionine methyl-ester for 1 h at room temperature. Membranes of equal protein content were incubated with indicated amounts of EboV GP_{ATM}, pre-treated or not with thermolysin, for 1 h at room temperature in the presence of Complete Protease Inhibitor Cocktail (Roche) and incubated for an additional hour on ice before the addition of membrane lysis buffer (12.5 mM CHAPSO, 150 mM NaCl, 1 mM EDTA, 10 mM Tris/HCl pH 7.4) for a final concentration of 10 mM CHAPSO. Proteins were solubilized on ice for 20 min and debris was removed by centrifugation at 12,000g for 10 min at 4 °C. The soluble membrane lysates were incubated with anti-NPC1 antibody for 1 h at 4 °C and then incubated with Protein A-agarose beads (Sigma) for an additional 4 h at 4 °C. Beads were then washed three times with 8 mM CHAPSO, 150 mM NaCl, 1 mM EDTA, 10 mM Tris/HCl pH 7.4 and immunoprecipitated product was eluted by incubation in 0.1 M glycine pH 3.5 for 5 min at room temperature. The eluted complex was then neutralized and analysed by immunoblot using the indicated antibody.

Photo-activation and click chemistry. Photo-activation and click chemistry were performed as described previously with some modifications²¹. Briefly, the 15,000g pellets from homogenized CHO_{hNPC1} or CHO_{null} cells were resuspended in PBS and incubated with the indicated concentrations of 3.47, 3.18 or DMSO for 10 min at room temperature. Membranes were then incubated with 25 µM of 3.98 for an additional 10 min and exposed to ultraviolet light (365 nm) for 1 min on ice. Proteins were solubilized in lysis buffer (1% Triton X-100, 0.1% NP-40, 20 mM HEPES pH 7.4) containing protease inhibitors and 150 µM of biotin-azide (Invitrogen) was added, followed by 5 mM L-ascorbic acid. The cycloaddition reaction (click chemistry) was initiated by the addition of 1 mM CuSO₄ and samples were incubated for 3 h at room temperature. NPC1 was immunoprecipitated and the product was resolved by SDS-PAGE, transferred to PVDF membrane, and analysed for conjugation of 3.98 to NPC1 using streptavidin-horseradish peroxidase (Sigma).

26. Gelsthorpe, M. E. *et al.* Niemann-Pick type C1 I1061T mutant encodes a functional protein that is selected for endoplasmic reticulum-associated degradation due to protein misfolding. *J. Biol. Chem.* **283**, 8229–8236 (2008).
27. Millard, E. E., Srivastava, K., Traub, L. M., Schaffer, J. E. & Ory, D. S. Niemann-pick type C1 (NPC1) overexpression alters cellular cholesterol homeostasis. *J. Biol. Chem.* **275**, 38445–38451 (2000).
28. Soneoka, Y. *et al.* A transient three-plasmid expression system for the production of high titer retroviral vectors. *Nucleic Acids Res.* **23**, 628–633 (1995).
29. Towner, J. S. *et al.* Generation of eGFP expressing recombinant Zaire ebolavirus for analysis of early pathogenesis events and high-throughput antiviral drug screening. *Virology* **332**, 20–27 (2005).
30. Weidmann, M., Mühlberger, E. & Hufert, F. T. Rapid detection protocol for filoviruses. *J. Clin. Virol.* **30**, 94–99 (2004).

A stress response pathway regulates DNA damage through β_2 -adrenoreceptors and β -arrestin-1

Makoto R. Hara¹, Jeffrey J. Kovacs¹, Erin J. Whalen¹, Sudarshan Rajagopal¹, Ryan T. Strachan¹, Wayne Grant², Aaron J. Towers^{1,3}, Barbara Williams¹, Christopher M. Lam¹, Kunhong Xiao¹, Sudha K. Shenoy¹, Simon G. Gregory^{1,3}, Seungkirl Ahn¹, Derek R. Duckett² & Robert J. Lefkowitz^{1,4}

The human mind and body respond to stress¹, a state of perceived threat to homeostasis, by activating the sympathetic nervous system and secreting the catecholamines adrenaline and noradrenaline in the 'fight-or-flight' response. The stress response is generally transient because its accompanying effects (for example, immunosuppression, growth inhibition and enhanced catabolism) can be harmful in the long term². When chronic, the stress response can be associated with disease symptoms such as peptic ulcers or cardiovascular disorders³, and epidemiological studies strongly indicate that chronic stress leads to DNA damage^{4,5}. This stress-induced DNA damage may promote ageing⁶, tumorigenesis^{4,7}, neuropsychiatric conditions^{8,9} and miscarriages¹⁰. However, the mechanisms by which these DNA-damage events occur in response to stress are unknown. The stress hormone adrenaline stimulates β_2 -adrenoreceptors that are expressed throughout the body, including in germline cells and zygotic embryos¹¹. Activated β_2 -adrenoreceptors promote Gs-protein-dependent activation of protein kinase A (PKA), followed by the recruitment of β -arrestins, which desensitize G-protein signalling and function as signal transducers in their own right¹². Here we elucidate a molecular mechanism by which β -adrenergic catecholamines, acting through both Gs-PKA and β -arrestin-mediated signalling pathways, trigger DNA damage and suppress p53 levels respectively, thus synergistically leading to the accumulation of DNA damage. In mice and in human cell lines, β -arrestin-1 (ARRB1), activated via β_2 -adrenoreceptors, facilitates AKT-mediated activation of MDM2 and also promotes MDM2 binding to, and degradation of, p53, by acting as a molecular scaffold. Catecholamine-induced DNA damage is abrogated in *Arrb1*-knockout (*Arrb1*^{-/-}) mice, which show preserved p53 levels in both the thymus, an organ that responds prominently to acute or chronic stress¹³, and in the testes, in which paternal stress may affect the offspring's genome. Our results highlight the emerging role of ARRB1 as an E3-ligase adaptor in the nucleus, and reveal how DNA damage may accumulate in response to chronic stress.

As a model of chronic stress and prolonged stimulation of β_2 -adrenoreceptors^{7,13}, wild-type mice were infused for four weeks with either saline or the β_2 -adrenoreceptor-agonist isoproterenol, a synthetic analogue of adrenaline. First, we tested whether this regimen affects DNA damage by examining phosphorylation of histone H2AX (γ -H2AX), one of the earliest indicators of DNA damage¹⁴. Isoproterenol infusion leads to DNA damage in the thymus (Fig. 1a, left panel). Accumulation of DNA damage indicates compromised genome maintenance. To investigate the potential mechanism, we examined p53 levels in the thymus and found that isoproterenol infusion leads to decreased levels of p53 (Fig. 1a, right panel). Consistent with the effects of isoproterenol *in vivo*, chronic stimulation of β_2 -adrenoreceptors with β -adrenergic catecholamines (isoproterenol, adrenaline or noradrenaline) leads to accumulation of DNA damage and a decrease in p53 levels in cultured U2OS cells (Supplementary

Fig. 1a–c), which endogenously express wild-type p53 and only the β_2 -subtype of β -adrenoreceptors (Supplementary Fig. 2a–c). Moreover, the p53 in these cells, as well as in all other cell lines used in these studies (fibroblasts and HEK-293 cells), was demonstrated to be functional by a variety of techniques (Supplementary Fig. 3a–k), and all cell lines endogenously expressed only the β_2 -subtype of β -adrenoreceptors (Supplementary Fig. 2a–c).

The isoproterenol-induced reduction in p53 levels results from p53 degradation, and is abolished by proteasome inhibition (Supplementary

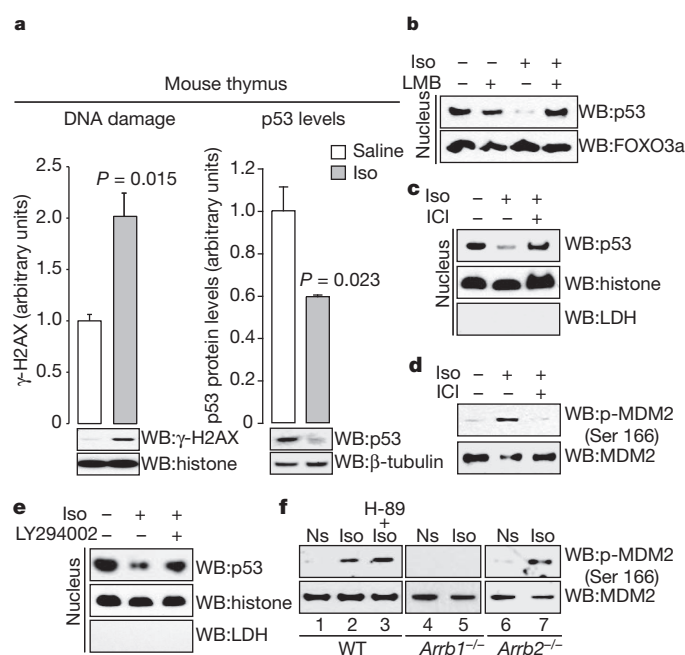


Figure 1 | Chronic catecholamine stimulation leads to p53 degradation and accumulation of DNA damage via ARRB1/AKT-mediated activation of MDM2. **a**, Isoproterenol infusion leads to accumulation of DNA damage and decreased p53 levels. Mice ($n = 3$ –5 for each condition) were infused with saline or isoproterenol ($30 \text{ mg kg}^{-1} \text{ d}^{-1}$) for 4 weeks. All bars represent mean \pm s.e.m. Histone, histone H2B; Iso, isoproterenol; WB, western blot. **b**, Isoproterenol-induced p53 reduction is dependent on nuclear export. This effect is specific to p53, in that another nuclear–cytosol shuttling molecule, FOXO3a, is not affected. LMB, leptomycin B. **c**, Preincubation with the β_2 -adrenoreceptor-selective antagonist ICI 118,551 (ICI) blocks isoproterenol-induced nuclear export of p53. Lactate dehydrogenase (LDH) is a cytosolic marker and histone is a nuclear marker. **d**, Isoproterenol stimulation leads to MDM2 phosphorylation at Ser 166, and is blocked by preincubation with ICI 118,551. **e**, Inhibition of the PI3K/AKT cascade abolishes isoproterenol-stimulated decreases in p53 levels in U2OS cells. LY294002 is a PI3K inhibitor. **f**, Isoproterenol stimulation leads to Gs-independent, ARRB1-dependent MDM2 phosphorylation at Ser 166. Ns, not stimulated.

¹Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710, USA. ²Translational Research Institute, The Scripps Research Institute, Jupiter, Florida 33458, USA. ³Center for Human Genetics, Duke University Medical Center, Durham, North Carolina 27710, USA. ⁴Howard Hughes Medical Institute, Duke University Medical Center, Durham, North Carolina 27710, USA.

Fig. 1d). Because nuclear export of p53 has been shown to be involved in its degradation¹⁵, we examined p53 localization. Subcellular fractionation shows that isoproterenol stimulation leads to a decrease in nuclear p53 and an increase in cytosolic p53 (Supplementary Fig. 1e, lower panels), thus, isoproterenol stimulation leads to p53 nuclear export. Immunocytochemical examination also shows increased levels of cytosolic p53 after isoproterenol stimulation (Supplementary Fig. 1e, upper panels). Isoproterenol concentrations as low as 1 nM lead to p53 nuclear export, resulting in a decrease in total p53 levels (Supplementary Fig. 1f). The importance of nuclear export in modulating p53 levels was investigated by treating cells with leptomycin B, an inhibitor of nuclear export. Leptomycin B pretreatment reverses isoproterenol-induced nuclear export of p53 (Fig. 1b).

To examine whether isoproterenol-induced effects were specifically mediated by β_2 -adrenoreceptors, U2OS cells were stimulated with isoproterenol in the presence or absence of the subtype-selective β_2 -adrenoreceptor antagonist ICI 118,551. Preincubation with ICI 118,551 abrogates the isoproterenol-induced decrease in p53 levels (Fig. 1c). During *in vivo* experiments, isoproterenol infusion leads to accumulation of DNA damage in the cerebellum, where β_2 -adrenoreceptors are the major subtype of β -adrenoreceptor¹⁶ (Supplementary Fig. 1g). Furthermore, targeted disruption of the *Adrb2* gene in mice markedly reduces accumulation of DNA damage upon isoproterenol infusion (Supplementary Fig. 1h). Taken together, these data indicate that stimulation of the β_2 -adrenoreceptor results in the nuclear export and degradation of p53 in a specific manner.

The E3 ligase MDM2 has been shown to have an important role in the regulation of p53 nuclear export and degradation¹⁵. Consistent with this, leptomycin B abrogates the ability of MDM2 to degrade p53 (ref. 15). Before MDM2-mediated ubiquitination of p53, the phosphoinositide 3-kinase (PI3K)/AKT cascade phosphorylates MDM2, activating its E3 ligase function¹⁷. To examine whether stimulation of β_2 -adrenoreceptors leads to MDM2 phosphorylation via the PI3K/AKT cascade, wild-type mouse embryonic fibroblasts (MEFs) were stimulated with isoproterenol in the presence or absence of ICI 118,551. Isoproterenol stimulation leads to MDM2 phosphorylation at Ser166, an AKT phosphorylation site, and the effect is

antagonized by ICI 118,551 (Fig. 1d and Supplementary Fig. 1i). To confirm that MDM2 is phosphorylated by the PI3K/AKT cascade upon isoproterenol stimulation, U2OS cells were stimulated with isoproterenol in the presence of either the PI3K inhibitor wortmannin or the AKT inhibitor AKTi. MDM2 phosphorylation is abolished by either wortmannin or AKTi (Supplementary Fig. 1j). Furthermore, a PI3K inhibitor also abolishes catecholamine-induced lowering of p53 levels in the nucleus (Fig. 1e and Supplementary Fig. 1k). The importance of MDM2 phosphorylation at Ser166 was demonstrated by the overexpression of a phosphomimetic mutant at Ser166 (MDM2-S166D)¹⁷, which facilitates the degradation of p53 when compared to wild-type MDM2 (Supplementary Fig. 1l). These data implicate the PI3K/AKT cascade downstream of the β_2 -adrenoreceptor as a mediator of p53 stability through the phosphorylation of MDM2.

Upon activation of β_2 -adrenoreceptors, the PI3K/AKT cascade can be stimulated by both the Gs-PKA¹⁸ and β -arrestin-mediated signalling pathways^{19,20}. To elucidate which pathway was involved, we examined the effect of β_2 -adrenoreceptor stimulation in wild-type MEFs in the presence of H-89, a PKA inhibitor, or in *Arrb1*^{-/-} or *Arrb2*-knockout (*Arrb2*^{-/-}) MEFs. In wild-type MEFs, H-89 does not inhibit isoproterenol-stimulated MDM2 phosphorylation (Fig. 1f, lane 3). In contrast, the isoproterenol effect is abrogated in *Arrb1*^{-/-} (Fig. 1f, lane 5), but not in *Arrb2*^{-/-} MEFs (Fig. 1f, lane 7). Furthermore, rescuing *Arrb1* expression in *Arrb1*^{-/-} MEFs restores the effects of isoproterenol stimulation (Supplementary Fig. 1m), and the effects of isoproterenol on MDM2 phosphorylation are abrogated in *Arrb1*^{-/-} mice (Supplementary Fig. 1n). These data elucidate a Gs-independent, ARRB1-dependent signalling pathway that regulates the activation state of MDM2 through the PI3K/AKT cascade.

β -Arrestins can serve as adaptors for E3 ligases and their substrates²¹. Because we found that MDM2 activation by the PI3K/AKT cascade is an ARRB1-dependent event, we examined the binding between β -arrestins and p53, a known MDM2 substrate. In HEK-293 cells stably overexpressing ARRB1 or ARRB2, p53 binds preferentially to ARRB1, an isoform localized to both the cytosol and nucleus²², but not to ARRB2, which predominantly localizes to the cytosol²²⁻²⁵ (Supplementary Fig. 4a). Binding between these two molecules at

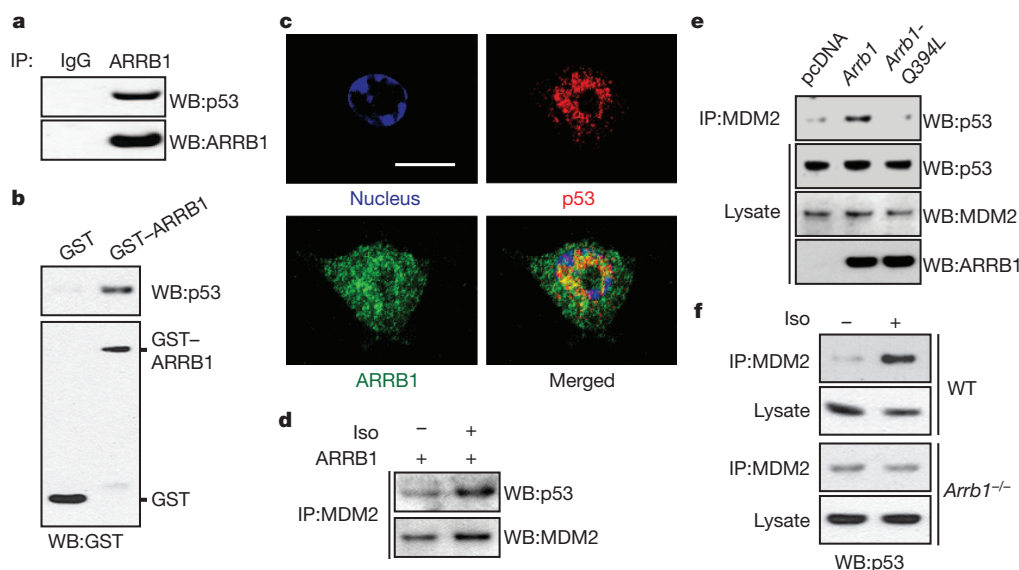


Figure 2 | ARRB1 functions as an E3 ligase adaptor for MDM2 and p53 upon catecholamine stimulation. **a**, Endogenous binding of p53 and ARRB1. Cell lysates from HEK-293 cells were used for immunoprecipitation (IP) with an anti-ARRB1 (K-16) antibody or normal IgG, and analysed by immunoblotting with anti-p53 (DO-1) antibody. **b**, *In vitro* binding of p53 and ARRB1. Purified p53 was incubated with either GST or GST-ARRB1 and precipitated with glutathione beads. Precipitates were analysed by immunoblotting with an anti-p53 (DO-1) antibody. **c**, Confocal analysis of

co-localization of p53 and ARRB1 in non-treated RAW264.7 macrophages, which endogenously express high levels of both p53 and ARRB1. Scale bar, 10 μ m. **d**, Isoproterenol stimulation facilitates the binding of MDM2 to p53 in ARRB1 overexpressed cells. **e**, Nuclear ARRB1 facilitates MDM2 binding to p53. U2OS cells were transfected with either empty vector (pcDNA), *Arrb1* or *Arrb1*-Q394L. **f**, ARRB1 facilitates isoproterenol-induced MDM2 binding to p53. WT, wild-type.

endogenous levels is also observed in HEK-293 cells and brain homogenates (Fig. 2a and Supplementary Fig. 4b). The binding seems to be direct, because purified ARRB1 tagged with glutathione-S-transferase (GST-ARRB1) binds to p53 *in vitro* (Fig. 2b). To examine the effect of β_2 -adrenoreceptor stimulation on this complex, we treated untransfected HEK-293 cells, which endogenously express only the β_2 subtype of β -adrenoreceptors (Supplementary Fig. 2b, c), with isoproterenol. Stimulation of β_2 -adrenoreceptors does not affect the binding of ARRB1 to p53 (Supplementary Fig. 4c). Subsequently, we mapped the binding sites in ARRB1 and p53 by performing sequential deletions, followed by immunoprecipitation from HEK-293 cells (Supplementary Fig. 4d, e). We identified the amino terminus of ARRB1 (amino acids 1–186) as critical for binding to p53. In p53, a domain comprising amino acids 101–186 is required for binding to ARRB1. Consistent with these results, a synthetic ARRB1-binding peptide (ARRB-BP), which binds to the N terminus of ARRB1 and induces a conformational change²⁶, disrupts the interaction between ARRB1 and p53 (Supplementary Fig. 4f).

Co-immunoprecipitation experiments after subcellular fractionation show that more than 90% of the binding between ARRB1 and p53 occurs in the nucleus (Supplementary Fig. 4g). Additionally, confocal analysis reveals that endogenous ARRB1 and p53 co-localize in the nucleus (Fig. 2c) and a ternary complex between ARRB1, MDM2 and p53 was observed in p53-null NCI-H1299 cells transfected with wild-type p53 (Supplementary Fig. 4h). The potential effects of nuclear ARRB1 on MDM2 binding to p53 were investigated in U2OS cells transfected with either *Arrb1* or *Arrb1-Q394L*, in which a single amino acid, Gln 394, has been mutated to Leu to create a nuclear export signal in ARRB1^{22,23}. Overexpression of ARRB1 facilitates the binding of MDM2 to p53, enhancing basal β_2 -adrenoreceptor-stimulated ARRB1 signalling (Fig. 2d); however, the effect is abolished with ARRB1-Q394L (Fig. 2e). This result indicates that ARRB1 facilitates an MDM2–p53 interaction in the nucleus. The role of endogenous ARRB1 as a facilitator of MDM2–p53 complex formation under isoproterenol-stimulated conditions is further demonstrated in *Arrb1*^{−/−} MEFs (Fig. 2f), in which loss of ARRB1 prevents the increased interaction of MDM2 and p53 after isoproterenol stimulation, when compared to wild-type cells.

Next we examined whether ARRB1 expression affects p53 levels by comparing different clonal populations of wild-type and *Arrb1*^{−/−} MEFs. *Arrb1*^{−/−} MEFs show increased p53 levels (Supplementary Fig. 5a). Furthermore, rescuing ARRB1 expression in *Arrb1*^{−/−} MEFs decreases p53 levels in a dose-dependent manner (Fig. 3a). Differences in p53 levels under basal conditions seem to be the result of decreased p53 ubiquitination in *Arrb1*^{−/−} MEFs (Fig. 3b, lanes 1 and 3). Furthermore, consistent with β_2 -adrenoreceptor-induced degradation of p53, isoproterenol stimulation promotes p53 ubiquitination, but the effect is markedly decreased in *Arrb1*^{−/−} MEFs (Fig. 3b, lanes 2 and 4). To address further whether ARRB1 facilitates the ubiquitination of p53 by MDM2, we conducted *in vitro* ubiquitination assays (Supplementary Fig. 5b). Addition of ARRB1 facilitates MDM2-mediated ubiquitination of p53 and the effect is abolished with ARRB-BP. Together, these data indicate that upon catecholamine stimulation, ARRB1 promotes the interaction of MDM2 and p53 by acting as an E3 ligase adaptor that facilitates ubiquitination of p53.

Cytosolic ARRB1 mediates catecholamine-induced activation of AKT and MDM2, whereas nuclear ARRB1 serves as an adaptor for MDM2-dependent ubiquitination of p53. Consistent with these results, isoproterenol stimulation leads to a lowering of p53 levels (Fig. 3c, lane 2). By contrast, p53 levels remain constant in *Arrb1*^{−/−} MEFs (Fig. 3c, lane 4). Furthermore, rescuing ARRB1 expression in *Arrb1*^{−/−} MEFs by transient transfection restores isoproterenol-stimulated degradation of p53 (Fig. 3c, lane 6), and suppression of ARRB1 by RNA interference results in suppression of isoproterenol-stimulated MDM2–p53 complex formation, and a lowering of p53 levels in U2OS cells (Fig. 3d, e).

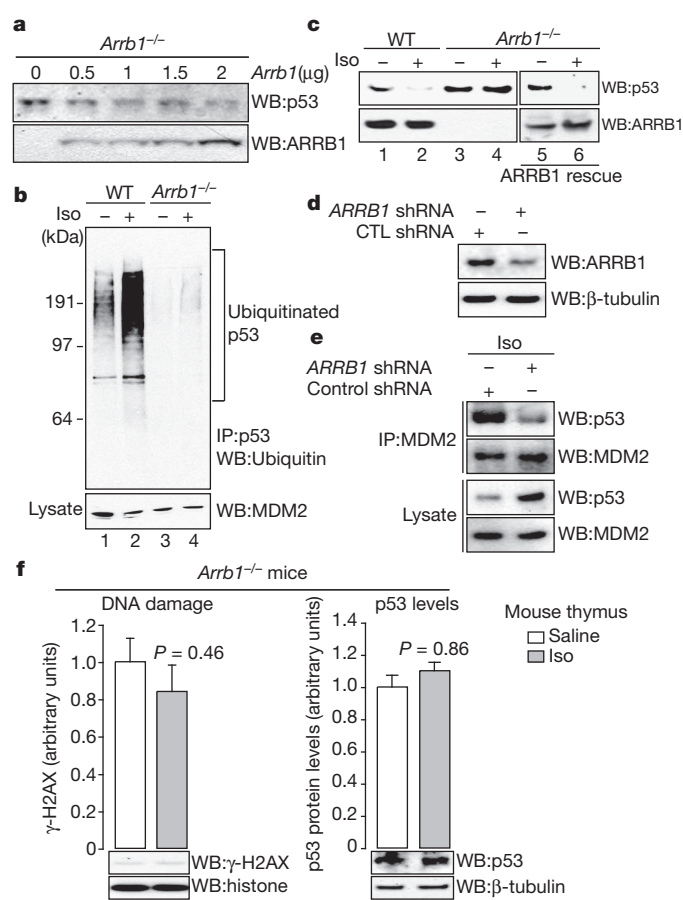


Figure 3 | ARRB1 facilitates catecholamine-induced p53 degradation by MDM2. **a**, Rescuing ARRB1 expression in *Arrb1*^{−/−} MEFs decreases p53 levels. *Arrb1*^{−/−} MEFs were transiently transfected with *Arrb1* and cell lysates were examined by immunoblotting. **b**, Isoproterenol stimulation leads to ubiquitination of p53 in an ARRB1-dependent manner. Wild-type and *Arrb1*^{−/−} MEFs were stimulated with 10 μ M isoproterenol for 24 h. Cell lysates were immunoprecipitated with an anti-p53 antibody (FL-393) and analysed by immunoblotting with an anti-ubiquitin (P4D1) antibody. **c**, Isoproterenol stimulation leads to ARRB1-dependent p53 degradation. **d**, Reduction in ARRB1 levels induced by small hairpin RNA (shRNA) in U2OS cells. **e**, Suppression of ARRB1 suppresses isoproterenol-induced binding of MDM2 to p53 and restores p53 levels in U2OS cells. U2OS cells were treated with either control shRNA or ARRB1 shRNA for 72 h, followed by 24 h stimulation with 10 μ M isoproterenol. **f**, Levels of p53 in isoproterenol-infused *Arrb1*^{−/−} mice remain constant and there is no accumulation of DNA damage. *Arrb1*^{−/−} mice ($n = 3$ –4 for each condition) were treated as in Fig. 1a. All bars represent mean \pm s.e.m.

To differentiate the cytosolic (catecholamine-induced MDM2 phosphorylation) and nuclear (E3 ligase adaptor) functions of ARRB1 in this cascade, a phosphomimetic mutant of MDM2 (*MDM2-S166D*) was co-transfected with either *Arrb1* or *Arrb1-Q394L* into *Arrb1*^{−/−} MEFs. This allowed us to focus on the nuclear function of ARRB1. Restoring ARRB1 expression with the wild type, but not with the Q394L mutant, facilitates the degradation of p53 (Supplementary Fig. 5c). Consequently, it seems that although the cytoplasmic pool of ARRB1 is sufficient to activate MDM2 through the PI3K/AKT pathway, the nuclear pool of ARRB1 is required to act as an E3 ligase adaptor for MDM2 towards p53.

To examine this cascade *in vivo*, we examined the effects of catecholamine on p53 levels and accumulation of DNA damage in the thymus of *Arrb1*^{−/−} mice. The mice were infused for four weeks with either saline or isoproterenol. In contrast to wild-type mice (Fig. 1a), p53 levels are maintained upon isoproterenol infusion in *Arrb1*^{−/−} mice, and accumulation of DNA damage is abrogated (Fig. 3f).

We have observed that isoproterenol infusion leads to lowering of p53 levels, and have elucidated a molecular mechanism whereby

ARRB1 regulates MDM2-dependent degradation of p53 upon β_2 -adrenoreceptor stimulation. Next, we investigated further the effects of catecholamine-dependent p53 degradation on accumulation of DNA damage. To visualize the prevalence of DNA damage, we analysed the formation of γ -H2AX foci in both wild-type and *Arrb1*^{-/-} MEFs. After chronic stimulation with isoproterenol, there is an increase in the formation of γ -H2AX foci (sevenfold) in wild-type MEFs, which is significantly reduced in *Arrb1*^{-/-} MEFs (Fig. 4a, panels 2 and 4, and Supplementary Fig. 6a). Moreover, rescuing ARRB1 expression in *Arrb1*^{-/-} MEFs restores the accumulation of isoproterenol-induced γ -H2AX foci (Fig. 4b, panels 5–8, and Supplementary Fig. 6a).

To examine how exposure to stress hormones initiates DNA damage, p53-null NCI-H1299 cells were chronically stimulated with isoproterenol. This leads to accumulation of DNA damage (Fig. 4c, panels 1–8), indicating that DNA damage is triggered by p53-independent mechanisms after isoproterenol stimulation. One of the prominent cascades leading to DNA damage is the generation of reactive oxygen species through Gs-PKA signalling²⁷. Accordingly, accumulation of isoproterenol-induced DNA damage is suppressed by inhibition of PKA (Fig. 4d, lanes 1 and 3). Consistent with the idea that ARRB1-mediated effects on DNA damage are due to altered p53 levels, rescuing p53 expression (Supplementary Fig. 3l) decreases isoproterenol-induced γ -H2AX foci (Fig. 4c, panels 9–16; Fig. 4d, lanes 1 and 2) and the p53 effect is antagonized by co-expression of ARRB1 (Fig. 4c, panels 17–20, and Supplementary Fig. 6b). These G-protein-mediated and ARRB1-mediated pathways may synergistically affect the accumulation of isoproterenol-induced DNA damage. Thus, combining PKA inhibition with rescue of p53 expression abrogates accumulation of DNA damage (Fig. 4d, lanes 1 and 4). Catecholamine-induced lowering of p53 levels may lead to increased survival of cells

containing DNA damage, owing to an impaired DNA damage checkpoint and repair cascade²⁸. This would then facilitate accumulation of DNA damage. Accordingly, U2OS cells were irradiated with ultraviolet light after isoproterenol stimulation. Chronic stimulation leads to increased FOS expression, an indicator of cell survival and proliferation (Supplementary Fig. 6c). Because DNA damage occurs under these conditions, FOS expression leads to proliferation of cells that contain DNA damage. Taken together, these data indicate that Gs-PKA-dependent signalling, which leads to the generation of reactive oxygen species²⁷, and ARRB1-dependent p53 degradation, which results in impaired DNA checkpoint and repair mechanisms²⁸, synergistically lead to accumulation of DNA damage and consequently may have effects on genomic integrity.

DNA damage may promote rearrangements in chromosomes. To quantify the occurrence of catecholamine-induced rearrangements, we analysed inter-chromosomal rearrangements between *Tcr γ* (the T-cell-receptor- γ locus) and *Tcr β* (the T-cell-receptor- β locus) in thymocytes (see Methods and Supplementary Fig. 6d). Both wild-type and *Arrb1*^{-/-} mice were infused for four weeks with either saline or isoproterenol, and genomic DNA was isolated from the thymus. Consistent with the accumulation of DNA damage (Figs 1a, 3f), isoproterenol infusion leads to an increase in these *Tcr* rearrangements in wild-type mice; however, the effects are no longer observed in *Arrb1*^{-/-} mice (Fig. 4e and Supplementary Fig. 6e). This indicates that catecholamine-stress-hormone-dependent accumulation of DNA damage promotes rearrangements in chromosomes.

Because chronic isoproterenol stimulation affects DNA damage and chromosomal rearrangements, we examined whether this cascade also affects genome integrity in the testes, in which paternal stress may affect the offspring's genome. Using the isoproterenol-infusion model of chronic stress^{7,13}, we observed that isoproterenol stimulation leads

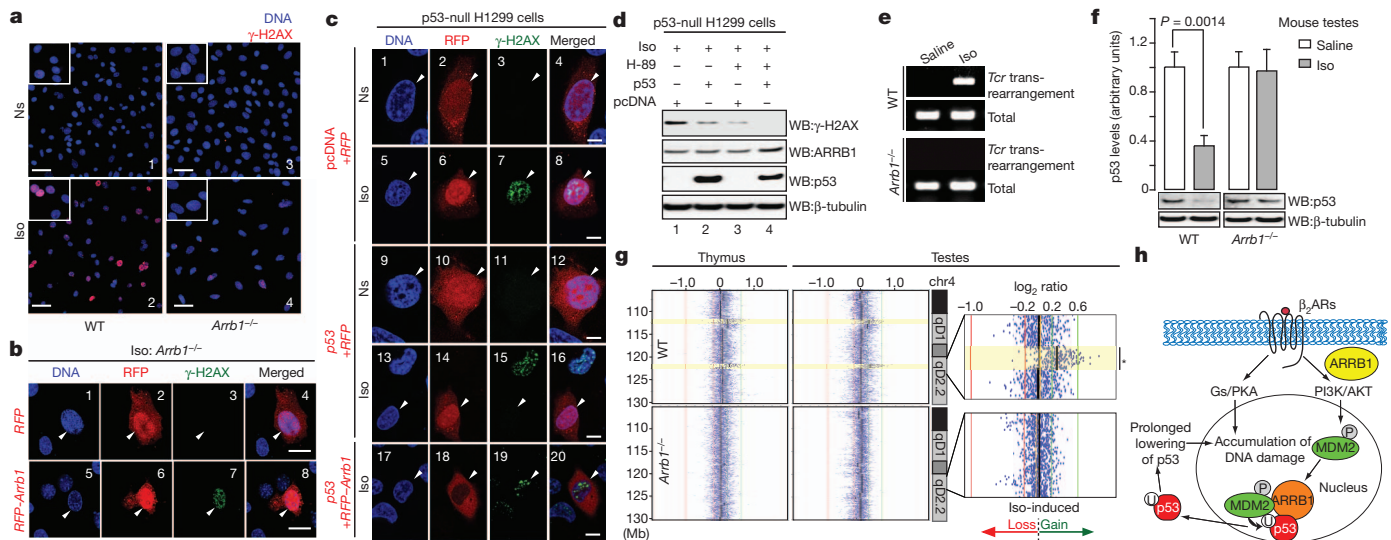


Figure 4 | Chronic catecholamine stimulation leads to accumulation of DNA damage by an ARRB1- and p53-dependent mechanism.

a, Isoproterenol stimulation leads to formation of γ -H2AX foci in wild-type but not in *Arrb1*^{-/-} MEFs. Wild-type and *Arrb1*^{-/-} MEFs were chronically stimulated with 10 μ M isoproterenol every 12 h for 3 days. Cells were immunostained and examined by confocal microscopy. Scale bar, 50 μ m. **b**, Rescuing ARRB1 expression in *Arrb1*^{-/-} MEFs restores isoproterenol-induced γ -H2AX foci. Two days after transfection, cells were stimulated and examined as described in **a**. RFP, red fluorescent protein. Scale bar, 10 μ m. **c**, ARRB1-dependent regulation of p53 levels mediates the accumulation of isoproterenol-induced DNA damage. Two days after the transfection, cells were stimulated as in **a**. Transfected cells, indicated with arrowheads, were visualized with RFP. Scale bar, 10 μ m. **d**, Isoproterenol-induced accumulation of DNA damage is synergistically suppressed by PKA inhibition and by rescuing p53 expression. **e**, Isoproterenol stimulated, ARRB1-dependent accumulation of DNA damage promotes rearrangements in chromosomes.

Total indicates PCR amplification of a non-specific locus (see Methods).

f, Isoproterenol infusion leads to decreased p53 levels in the testes from wild-type, but not *Arrb1*^{-/-}, mice. Wild-type and *Arrb1*^{-/-} mice ($n = 5$ for each condition) were treated as in Fig. 1a. All bars represent mean \pm s.e.m.

g, Isoproterenol-infused mice develop chromosomal rearrangements in an ARRB1-dependent manner. Wild-type and *Arrb1*^{-/-} mice were infused as in Fig. 1a. Genomic DNA from each organ was examined in an array-CGH. The data represent log₂ ratio plots (isoproterenol/saline) of genomic content in chromosome 4 (chr4, 105–130 Mb), comparing isoproterenol-infused mice with saline-infused mice of a same genotype. *, significance threshold of 1.0×10^{-7} (rank segmentation algorithm). The direction of isoproterenol-induced chromosomal gain (green arrow) or loss (red arrow) is indicated. Yellow highlights represent sites of isoproterenol-induced rearrangements. **h**, Schematic diagram of β_2 -adrenoreceptor (β_2 AR)-dependent regulation of DNA damage in response to prolonged secretion of catecholamines during chronic stress.

to a lowering of p53 levels in the testes. The effects are abolished in *Arrb1*^{-/-} mice (Fig. 4f). To examine these phenomena in a genome-wide context, we conducted an array-comparative genomic hybridization (array-CGH). In the same model of chronic stress^{7,13}, genomic DNA was isolated from the testes and thymus, which allowed us to eliminate any changes due to meiotic recombination by considering only rearrangements that occurred in both organs (see study design in Supplementary Fig. 6f). These studies show that the only such rearrangement occurring upon isoproterenol-infusion results in a duplication of more than 1 megabase (Mb) in regions 4qD2.2 and 4qD1 in wild-type mice; however, these events are not observed in the testes from *Arrb1*^{-/-} mice (Fig. 4g and Supplementary Fig. 6f). Quantitative PCR (qPCR) of the testicular genome from each mouse also confirms isoproterenol-induced duplication at 4qD2.2 in an *ARRB1*-dependent manner (Supplementary Fig. 6g). Taken together, these data support the hypothesis that β_2 -adrenoreceptor- and *ARRB1*-dependent signalling regulates catecholamine-induced degradation of p53, thus leading to the accumulation of DNA damage in both somatic and germline cells (Fig. 4h).

The stress response is conserved in mammals, and is probably required for survival. However, psychosocial stress in humans is not time-limited, because aspects of this type of stress response can be sustained over months or even years. This may lead to prolonged secretion of stress hormones and consequent adverse effects for the individual. Indeed, clinical studies have shown marked risk-reductions for prostate cancer, lung adenocarcinoma and Alzheimer's disease associated with chronic β -blocker (β -adrenoreceptor-antagonist) therapy^{4,29,30}. It also seems plausible that such hormonal influences on DNA damage may not be limited to the β_2 -adrenoreceptors.

METHODS SUMMARY

Experimental procedures. Each experiment was repeated at least three times with comparable results, unless indicated otherwise.

Cell culture conditions and treatments. Isoproterenol was prepared fresh for each experiment by dissolving bitartrate salt (Sigma) immediately before stimulation. To study chronic β -adrenergic effects, U2OS cells and MEFs were cultured until confluent, then stimulated with 10 μ M isoproterenol for 24 h unless otherwise indicated. To study γ -H2AX formation, cells were cultured until 40–50% confluent, then stimulated with 10 μ M isoproterenol every 12 h for 3 days. To study phosphorylation of MDM2 at Ser 166 in MEFs, cells were serum-starved for 4 h, then stimulated with 10 μ M isoproterenol for 1 h. H-89 (10 μ M), leptomycin B (10 nM), ICI 118,551 (10 μ M), wortmannin (100 nM), 5-(2-benzothiazolyl)-3-ethyl-2-(2-(methylphenylamino)ethyl)-1-phenyl-1H-benzimidazolium iodide (AKT1, 1 μ M) or LY294002 (10 μ M, Sigma) were added to the media 30 min before stimulation with isoproterenol.

Isoproterenol infusion. Mice were subcutaneously implanted with ALZET osmotic pumps to administer saline or isoproterenol (30 mg kg⁻¹ d⁻¹) continuously, dissolved in saline, for 28 days (mini-osmotic pump model 2004), following the manufacturer's procedure. After administration, animals were killed and the indicated organs were dissected out. All animals used in these studies were adult male mice of 8–12 weeks of age. Animals were handled according to approved protocols and animal welfare regulations of the Institutional Review Board at Duke University Medical Center.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 July 2010; accepted 18 July 2011.

Published online 21 August 2011.

1. Selye, H. A syndrome produced by diverse nocuous agents. *Nature* **138**, 32 (1936).
2. Charmandari, E., Tsigos, C. & Chrousos, G. Endocrinology of the stress response. *Annu. Rev. Physiol.* **67**, 259–284 (2005).
3. Goldstein, D. S. Catecholamines and stress. *Endocr. Regul.* **37**, 69–80 (2003).
4. Antoni, M. H. et al. The influence of bio-behavioural factors on tumour biology: pathways and mechanisms. *Natl. Rev.* **6**, 240–248 (2006).
5. Flint, M. S., Baum, A., Chambers, W. H. & Jenkins, F. J. Induction of DNA damage, alteration of DNA repair and transcriptional activation by stress hormones. *Psychoneuroendocrinology* **32**, 470–479 (2007).

6. Lu, T. et al. Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883–891 (2004).
7. Thaker, P. H. et al. Chronic stress promotes tumor growth and angiogenesis in a mouse model of ovarian carcinoma. *Nature Med.* **12**, 939–944 (2006).
8. Fratiglioni, L., Paillard-Borg, S. & Winblad, B. An active and socially integrated lifestyle in late life might protect against dementia. *Lancet Neurol.* **3**, 343–353 (2004).
9. Kinney, D. K., Munir, K. M., Crowley, D. J. & Miller, A. M. Prenatal stress and risk for autism. *Neurosci. Biobehav. Rev.* **32**, 1519–1532 (2008).
10. Nepomnaschy, P. A. et al. Cortisol levels and very early pregnancy loss in humans. *Proc. Natl Acad. Sci. USA* **103**, 3938–3942 (2006).
11. Ćikoš, S. et al. Expression of beta adrenergic receptors in mouse oocytes and preimplantation embryos. *Mol. Reprod. Dev.* **71**, 145–153 (2005).
12. DeWire, S. M., Ahn, S., Lefkowitz, R. J. & Shenoy, S. K. β -arrestins and cell signaling. *Annu. Rev. Physiol.* **69**, 483–510 (2007).
13. Ni, Y. et al. Activation of β_2 -adrenergic receptor stimulates γ -secretase activity and accelerates amyloid plaque formation. *Nature Med.* **12**, 1390–1396 (2006).
14. Bonner, W. M. et al. GammaH2AX and cancer. *Nature Rev. Cancer* **8**, 957–67 (2008).
15. Freedman, D. A. & Levine, A. J. Nuclear export is required for degradation of endogenous p53 by MDM2 and human papillomavirus E6. *Mol. Cell. Biol.* **18**, 7288–7293 (1998).
16. Rainbow, T. C., Parsons, B. & Wolfe, B. B. Quantitative autoradiography of beta 1- and beta 2-adrenergic receptors in rat brain. *Proc. Natl Acad. Sci. USA* **81**, 1585–1589 (1984).
17. Zhou, B. P. et al. *HER-2/neu* induces p53 ubiquitination via Akt-mediated MDM2 phosphorylation. *Nature Cell Biol.* **3**, 973–982 (2001).
18. De Gregorio, G. et al. The p85 regulatory subunit of PI3K mediates TSH-cAMP-PKA growth and survival signals. *Oncogene* **26**, 2039–2047 (2007).
19. Goel, R., Phillips-Mason, P. J., Raben, D. M. & Baldassare, J. J. α -Thrombin induces rapid and sustained Akt phosphorylation by β -arrestin1-dependent and -independent mechanisms, and only the sustained Akt phosphorylation is essential for G1 phase progression. *J. Biol. Chem.* **277**, 18640–18648 (2002).
20. Buchanan, F. G. et al. Role of β -arrestin 1 in the metastatic progression of colorectal cancer. *Proc. Natl Acad. Sci. USA* **103**, 1492–1497 (2006).
21. Kovacs, J. J., Hara, M. R., Davenport, C. L., Kim, J. & Lefkowitz, R. J. Arrestin development: emerging roles for β -arrestins in developmental signaling pathways. *Dev. Cell* **17**, 443–458 (2009).
22. Scott, M. G. et al. Differential nucleocytoplasmic shuttling of β -arrestins. Characterization of a leucine-rich nuclear export signal in β -arrestin-2. *J. Biol. Chem.* **277**, 37693–37701 (2002).
23. Wang, P., Wu, Y., Ge, X., Ma, L. & Pei, G. Subcellular localization of β -arrestins is determined by their intact N domain and the nuclear export signal at the C terminus. *J. Biol. Chem.* **278**, 11648–11653 (2003).
24. Wang, P. et al. β -arrestin 2 functions as a G-protein-coupled receptor-activated regulator of oncoprotein Mdm2. *J. Biol. Chem.* **278**, 6363–6370 (2003).
25. Boularan, C. et al. β -arrestin 2 oligomerization controls the Mdm2-dependent inhibition of p53. *Proc. Natl Acad. Sci. USA* **104**, 18061–18066 (2007).
26. Nobles, K. N., Guan, Z., Xiao, K., Oas, T. G. & Lefkowitz, R. J. The active conformation of β -arrestin1: direct evidence for the phosphate sensor in the N-domain and conformational differences in the active states of β -arrestins1 and -2. *J. Biol. Chem.* **282**, 21370–21381 (2007).
27. Yan, L. et al. Type 5 adenylyl cyclase disruption increases longevity and protects against stress. *Cell* **130**, 247–258 (2007).
28. Sengupta, S. & Harris, C. C. p53: traffic cop at the crossroads of DNA repair and recombination. *Nature Rev. Mol. Cell Biol.* **6**, 44–55 (2005).
29. Schuller, H. M. Mechanisms of smoking-related lung and pancreatic adenocarcinoma development. *Nature Rev. Cancer* **2**, 455–463 (2002).
30. Khachaturian, A. S. et al. Antihypertensive medication use and incident Alzheimer disease: the Cache County Study. *Arch. Neurol.* **63**, 686–692 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements R.J.L. is a Howard Hughes Medical Institute investigator. This work was supported by HL16037 and HL70631 (R.J.L.). We thank D. Addison and Q. Lennon for secretarial assistance; S. H. Snyder and M. A. Koldobsky for providing p53 deletion constructs; B. K. Kobilka and H. A. Rockman for providing *Adrb2*^{-/-} mice; M. C. Hung for providing the MDM2-S166D plasmid; A. K. Shukla, A. Kahsai, J. Kim, J. Sun, S. M. DeWire and N. Odajima for discussion and comments.

Author Contributions M.R.H. and R.J.L. designed experiments, directed the study and wrote the paper. M.R.H. performed most of the experiments, analysed the data and prepared the figures. R.J.L. supervised the study and provided financial support. J.J.K. performed some experiments, helped to analyse the data and helped to write the paper. E.J.W., S.R., K.X., S.K.S. and S.A. helped to analyse the data. E.J.W., S.R., R.T.S., B.W., C.M.L. and S.A. performed some experiments. A.J.T. and S.G.G. performed the array-CGH and helped to analyse the data. W.G. and D.R.D. helped to characterize the functionality of p53 and to analyse the data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.J.L. (lefko001@receptor-biol.duke.edu).

METHODS

Reagents. Unless otherwise noted, chemicals were purchased from Sigma.

Antibodies. Antibodies used were as follows, indicated WB for western blotting, IP for immunoprecipitation and CM for confocal microscopy. PARP-1 (WB, 1:500 dilution, Alexis Biochemicals). FOXO3a (WB, 1:500), MDM2 phosphorylated at Ser 166 (WB, 1:1,000), FOS (WB, 1:500), PUMA (WB, 1:500), p21 (WB, 1:500), all from Cell Signaling. Mouse p53 (FL-393; IP, 2 µg; WB, 1:200; CM, 1:50), human p53 (DO-1; WB, 1:5,000), MDM2 (SMP14; IP, 1 µg), ARRB1 (K-16; IP, 1 µg), ubiquitin (P4D1; WB, 1:200), human β_2 -adrenoreceptor (H-20; WB, 1:3,000), mouse β_2 -adrenoreceptor (M-20; WB, 1:200), anti-goat IgG-HRP (WB, 1:10,000), all from Santa Cruz. ARRB1 (10; WB, 1:200; CM, 1:50; BD Biosciences). LDH (WB, 1:300; Calbiochem). Mdm2 (HDM2-323; WB, 1:200), β -tubulin I (SAP4G5; WB, 1:10,000), both from Sigma. Histone H2B (WB, 1:5,000), histone H2B phosphorylated on Ser 14 (WB, 1:5,000), γ -H2AX (WB, 1:1,000; CM, 1:100), all from Millipore. 53BP1 (WB, 1:1,000; Novus). p21 (WB, 1:200; Rockland). Anti-mouse IgG-HRP (WB, 1:10,000), anti-rabbit IgG-HRP (WB, 1:10,000), both from GE Healthcare. Rabbit polyclonal ARRB1 antibody (A1CT; WB, 1:20,000) was generated as previously described³¹.

Primers. *Tcrp* a1: 5'-ACCATACACTGGTACCGGCA-3', *Tcrp* b1: 5'-ACCCC TACCCATATTTCTTAG-3', *Tcrb* a2: 5'-TCTACTCCAACTACTCCAG-3', *Tcrb* b2: 5'-CCTCCAAGCGAGGAGATGTGAA-3', non-specific locus (chr 7) forward: 5'-AGGCCTGGCTAGGCTTTTGAATCTTTC-3', non-specific locus (chr 7) reverse: 5'-TGCCAGTGCTGGTGCCTGTGCACGGCTGT-3', qPCR chr4 qD2.2 forward: 5'-TGGTGGCTGGCACAACCTGGCA-3', qPCR chr4 qD2.2 reverse: 5'-TGACGGTGTCTTTTGCCTTACAGAAGC-3', qPCR control (chr1) forward: 5'-CCTCCCATCAACGTTTCAGGAGCC-3', qPCR control (chr1) reverse: 5'-ACTGCTTCTGCTCCAAACCTGC-3', *p21* promoter forward: 5'-CCAGAGGATACCTTGAAGGC-3', *p21* promoter reverse: 5'-TCTCTGT CTCCATTCATGCTCTCC-3'.

Peptides. The synthesis of ARRB1-binding peptide (ARRB1-BP; V₂Rpp) has been described elsewhere. The sequence of the peptide, with phosphorylation sites underlined, is: ARGTRPPSLGPQDESCTTASSSLAKDTSS (ref. 32).

Plasmids. The MDM2-S166D plasmid¹⁷ and *p53* deletion constructs³³ and were gifts from M. C. Hung and S. H. Snyder, respectively. Plasmids encoding shRNAs against *p53* (psiRNA-mp53 and psiRNA-hp53), and the control psiRNA-LucGL3, were purchased from InvivoGen.

Experimental procedures. Each experiment was repeated at least three times with comparable results, unless indicated otherwise.

Immunoblotting. SDS polyacrylamide gel electrophoresis (SDS-PAGE) was performed on 1.0-mm-thick NuPAGE 4–12% Bis-Tris gels (Invitrogen) and separated proteins were transferred to nitrocellulose membranes by semi-dry transfer, using trans-blot transfer medium (Bio-Rad). Blots were blocked with blocking buffer (5% skimmed milk in PBS with 0.02% Tween-20) before incubation at 4 °C overnight with primary antibodies, diluted in blocking buffer as described above. Blots were washed three times for 5 min each in PBS with 0.02% Tween-20, and then incubated with secondary antibodies in blocking buffer. Blots were washed three times for 5 min each in PBS with 0.02% Tween-20, and developed by SuperSignal West Pico/Femto solution (Pierce). Each protein band of interest on the immunoblot was quantified by densitometry using the GeneTools program (SynGene).

Co-immunoprecipitation. Cells were lysed in a lysis buffer (50 mM Tris (pH 7.4), 150 mM NaCl, 0.1% CHAPS buffer, 0.1 mg ml⁻¹ BSA, 1 mM PMSF and 1 mM EDTA, with Halt protease and phosphatase inhibitor cocktail (Pierce)), and homogenized by passing through a 28-gauge needle 20 times. Crude lysates were cleared of insoluble debris by centrifugation at 14,000g. Extra lysis buffer was added to 100–500 µg of cell lysate to bring samples to a total volume of 1 ml. Immunoprecipitating antibody (1–2 µg) was added and incubated on a rotator at 4 °C overnight. On the following day, 25 µl (50% slurry) of the appropriate TrueBlot IP beads (eBioscience) was added and incubated on a rotator at 4 °C for 1 h. The beads were washed five times with the lysis buffer and quenched with 30 µl of SDS sample buffer (×2). For detection of p53 ubiquitination (Fig. 3b), 10 mM *N*-ethylmaleimide and 20 µM MG132 were added to the lysis buffer. Co-immunoprecipitation after cell fractionation was conducted as previously described³⁴. Briefly, cells were lysed in RIPA A buffer (0.3% Triton X-100, 50 mM Tris (pH 7.4) and 1 mM EDTA), with rotation at 4 °C for 30 min. Cell lysates were centrifuged at 14,000g for 10 min and the supernatant was used as the cytosolic fraction. The nuclear fraction was extracted from the pellet with RIPA B buffer (1% Triton X-100, 1% SDS, 50 mM Tris (pH 7.4), 500 mM NaCl and 1 mM EDTA), affinity-precipitated with the indicated antibodies, and subjected to SDS-PAGE.

Subcellular fractionation. U2OS cells from a 10-cm plate were resuspended in 300 µl of buffer B (0.25 M sucrose, 10 mM Tris (pH 7.4), 10 mM MgCl₂, 10 mM KCl, 1 mM DTT and protease inhibitor cocktail without EDTA) and homogenized with 150 strokes in a 1-ml dounce tissue grinder (Wheaton) using a tight pestle on

ice. After centrifugation at 750g for 10 min, the supernatant was isolated to separate the cytosolic fraction. The pellet was washed twice with buffer B, and resuspended in 100–200 µl of buffer B. The suspension was analysed as the nuclear fraction. Cytosolic fractions and nuclei were also prepared by using Nuclei EZ prep nuclei isolation kit (Sigma), following the manufacturer's protocol, and comparable results were obtained. To detect the effects of PI3K inhibition by LY294002 on isoproterenol-induced p53 nuclear export, U2OS cells were pre-incubated with 10 µM LY294002 for 30 min, and then stimulated with 10 µM isoproterenol for 1 h. To detect the effects of nuclear export on decreased nuclear p53, U2OS cells were pre-incubated with 10 nM leptomycin B for 30 min, and then stimulated with 10 µM isoproterenol for 1 h. To prepare a total-cell extract, cell pellets were lysed in a lysis buffer (50 mM Tris (pH 7.4), 150 mM NaCl, 0.1% CHAPS, 0.1 mg ml⁻¹ BSA, 1 mM PMSF and 1 mM EDTA, with Halt protease and phosphatase inhibitor cocktail (Pierce)) and homogenized by passing through a 28-gauge needle 20 times. Crude lysates were cleared of insoluble debris by centrifugation at 20,000g.

Cell culture conditions and treatments. Wild-type MEFs (passage number ~72), *Arrb1*^{-/-} MEFs (passage number ~76) and *Arrb2*^{-/-} MEFs (passage number ~49) were prepared according to the 3T3 protocol^{35,36}. Established MEF cultures and RAW264.7 cells were maintained in Dulbecco's modified Eagle medium (DMEM) with 10% FBS and 2 mM L-glutamine at 37 °C with a 5% CO₂ atmosphere in a humidified incubator. U2OS, HEK-293 and NCI-H1299 cells were maintained in modified Eagle medium (MEM) with 10% FBS and 2 mM L-glutamine, with the same conditions as above. U2OS and NCI-H1299 cells were transfected with FuGENE6 transfection reagent (Roche) following the manufacturer's protocol. For RNA interference for ARRB1, the vector system shRNA was used as previously described^{37,38}. Briefly, U2OS cells in 10-cm plates were transfected with either 10 µg control shRNA plasmid (5'-ACGTGACACGTTCCGAGAATTGATATCCGTTTC TCCGAACGTGTACAGTTT-3') or 10 µg *ARRB1* shRNA plasmid (5'-ATTCT CCGCGCAGAAGGCTTT GATATCCG AGCCTTCTGCGCGGAGAATTT-3'), and incubated for 72 h. HEK-293 cells and MEFs were transfected with lipofectamine 2000 (Invitrogen) following the manufacturer's protocol. Briefly, 2 µg of DNA was dissolved in 35 µl of serum- and antibiotic-free medium per well, in a 6-well plate. Lipofectamine 2000 (10 µl) was mixed with 25 µl of serum- and antibiotic-free medium, and incubated for 5 min. The prepared DNA and lipofectamine 2000 solutions were mixed and the mixture was incubated for 20 min at 18–23 °C. During the incubation, normal cell-culture medium was replaced with serum- and antibiotic-free medium. The transfection mixture was added to the cells in serum-free culture, and incubated overnight. On the following day, the medium was replaced with normal serum- and antibiotic-containing growth medium, and the cells were incubated for 48–72 h before testing.

Isoproterenol, epinephrine and norepinephrine were prepared fresh for each experiment by dissolving the bitartrate salts (Sigma) immediately before stimulation. To study chronic β -adrenergic effects, U2OS cells and MEFs were cultured until confluent, then stimulated with 10 µM isoproterenol for 24 h, unless otherwise indicated. To study γ -H2AX formation, cells were cultured until 40–50% confluent, then stimulated with 10 µM isoproterenol every 12 h for 3 days. H-89 (10 µM), leptomycin B (10 nM), ICI 118,551 (100 nM), wortmannin (100 nM), 5-(2-benzothiazolyl)-3-ethyl-2-(2-(methylphenylamino)ethyl)-1-phenyl-1H-benzimidazolium iodide (AKTi, 1 µM) or LY294002 (10 µM, Sigma) were added to the media 30 min before stimulation with isoproterenol. To study the effects of isoproterenol stimulation on cell proliferation after DNA damage, cells were ultraviolet-irradiated (50 J per m²) and incubated for 6 h, followed by stimulation with 10 µM isoproterenol every 12 h for 3 days. Cell lysates were examined by immunoblotting for FOS, an indicator of cell survival and proliferation³⁹. To study phosphorylation of MDM2 at Ser 166 in MEFs, cells were serum-starved for 4 h, then stimulated with 10 µM isoproterenol for 1 h. To study this phosphorylation event in U2OS cells, cells were serum-starved for 36 h, then stimulated with 10 µM isoproterenol for 10 min.

In vitro ubiquitination assay. 10 nM His-p53 (ProteinOne) was mixed with 200 ng E1 (BostonBiochem), 200 ng UbcH5b (BostonBiochem), 5 µg ubiquitin (BostonBiochem) and 25 nM MDM2 in 20 µl of reaction mixture (40 mM Tris (pH 7.6), 2 mM ATP-Mg²⁺, 1 mM dithiothreitol and 5 mM MgCl₂). Purified recombinant ARRB1 (0, 50 or 500 nM) was added to the reaction mixture in the presence or absence of 300 nM ARRB1-BP. The sample was incubated for 60 min at 30 °C, resolved by SDS-PAGE and analysed by immunoblotting with anti-p53 antibody (DO-1).

Isoproterenol infusion. Wild-type (C57BL/6), *Arrb1* knockout (*Arrb1*^{-/-})⁴⁰ or β_2 -adrenoreceptor knockout (*Adrb2*^{-/-})⁴¹ mice were subcutaneously implanted with ALZET osmotic pumps to administer saline or isoproterenol (30 mg kg⁻¹ d⁻¹) continuously, dissolved in saline, for 28 days (mini-osmotic pump model 2004), following the manufacturer's procedure. After administration, animals were killed and the indicated organs were dissected out. For protein preparation, dissected organ tissues were lysed and sonicated in RIPA buffer (50 mM Tris

(pH 7.4), 500 mM NaCl, 1% SDS, 1% Triton X-100 and 1 mM EDTA, with Halt protease and phosphatase inhibitor cocktail). Genomic DNA was prepared from dissected organ tissues by DNeasy blood & tissue kit (Qiagen), following the manufacturer's protocol. All animals used in these studies were adult male mice of 8–12 weeks of age. All mouse strains were backcrossed to the C57BL/6 background for ≥ 10 generations. Animals were handled according to approved protocols and animal welfare regulations of the Institutional Review Board at Duke University Medical Center.

Quantitative real-time PCR (qPCR). qPCR was performed with Power SYBR Green PCR Master Mix (Applied Biosystems) and StepOne Real-time PCR system (Applied Biosystems) following the manufacturer's protocol. To validate the array-CGH analysis, relative genomic content (copy number) was determined with the comparative C_T ($\Delta\Delta C_T$) method⁴².

GST pulldown assay. Wild-type rat *Arrb1* or human *MDM2* were subcloned into the pGEX4T1 vector and prepared according to the manufacturer's recommendations (Amersham Biosciences). The GST tag was cleaved with thrombin protease (Hematologic Technologies Inc.). p53 (1 nM) was co-incubated overnight with 10 nM of GST-ARRB1 or 10 nM of GST at 4 °C in 1 ml binding buffer (50 mM Tris (pH 7.4), 150 mM NaCl, 0.1 mg ml⁻¹ BSA and 10 μ M D-myo-inositol 1,2,3,4,5,6-hexakisphosphate), and 20 μ l of 50% glutathione-sepharose was then added to the mixture. The mixture was further incubated at 4 °C for 1 h with rotation. The beads were washed once with 1 ml binding buffer, separated by SDS-PAGE and analysed by immunoblotting.

Immunofluorescence experiments. Immunofluorescence using confocal microscopy was carried out as previously described⁴³. For detection of γ -H2AX foci, we captured images of more than 20 fields per preparation, which were randomly chosen in a blind manner. Cells positive for γ -H2AX foci in each field were tallied and added together to determine the percentage. The total number of cells was counted with 4',6-diamidino-2-phenylindole (DAPI) nuclear staining. For rescue experiments using the expression of RFP-ARRB1 (or RFP as a control) in MEFs, RFP-positive cells were counted for γ -H2AX foci. For p53 rescue experiments in NC1-H1299 cells, RFP-*Arrb1* (or RFP) and p53 were co-transfected in a 1:3 ratio.

Detection of interchromosomal rearrangements between *Tcrg* and *Tcrb*. The trans-rearrangement between *Tcrg* and *Tcrb* loci were detected by nested PCR, using first the 'a' set of primers and then the 'b' set of primers, as previously described⁴⁴ (Supplementary Fig. 6d). The number of rearrangements is expressed as the reciprocal of the highest dilution of DNA yielding an amplified product (for example, the number of trans-rearrangements per 1.5×10^5 cells (1 μ g of DNA) is 1,000, yielding an amplifiable fragment at a 1:1,000 dilution (1 ng of DNA))⁴⁴.

DNA preparation. DNA was prepared from the testes and thymus by using the DNeasy blood & tissue kit (Qiagen) following the manufacturer's protocol. To enrich sperm from excised testis grafts, the testis was minced and the epithelial tissue, containing leydig and sertoli cells, was removed.

Array-comparative genomic hybridization (Array-CGH). A tiling-path CGH array for the genome analysis in mouse (UCSC Build mm9) was designed and constructed by NimbleGen Systems (NimbleGen). The resulting array contained 720,000 probes with a median probe spacing of 3,537 base pairs. Probes were synthesized using an isothermal format (melting temperature 76 °C), and varied in length from 50 to 75 base pairs. Genomic DNA from five mice for each experimental condition was pooled and examined. Genomic DNAs (1 μ g) from test (isoproterenol-treated) and reference (saline-treated) mice were differentially labelled with 5'-Cy3 and 5'-Cy5 random nonamers (TriLink Biotechnologies), respectively, and hybridized to the oligoarray for 72 h using the MAUI hybridization station (BioMicro Systems Inc.). Image-capture of the hybridized arrays for fluorescent intensity extraction was performed using a Genepix 4100A scanner (Molecular Dynamics) and normalized using NimbleScan v2.5 microarray software (Nimblegen) before importing into Nexus Copy-Number (BioDiscovery) for analysis.

Array-CGH analysis. BioDiscovery's rank segmentation algorithm, which is similar to circular binary segmentation⁴⁵, was used to identify genomic rearrangements. The significance threshold was set as 1.0×10^{-7} . The calling algorithm used cluster values and defined \log_2 thresholds of ± 0.2 . We applied a cutoff of ten oligomer clones showing the same trend in copy-number change to define chromosomal rearrangements. Black lines in the plot indicate a 'cluster value', which is the median log-ratio value of all the probes in that region. Isoproterenol-induced rearrangements in the testes were determined, and identical rearrangements were detected in the thymus (see study design in Supplementary Fig. 6f).

Radioligand binding experiments. For ICI 118,551 and CGP 20712A affinity measurements, subtype-selective ligand affinities were determined from competition radioligand binding experiments, conducted according to previous works^{46,47}. Briefly, 25 μ g of cell membranes, prepared via differential centrifugation, were resuspended in assay buffer (50 mM Tris-HCl (pH 7.4), 12.5 mM MgCl₂, 2 mM

EDTA and 1 mM ascorbic acid) containing 60 pM [¹²⁵I]cyanopindolol (NEX189, 2,200 Ci mmol⁻¹) and concentrations of ICI 118,551 or CGP 20712A ranging from 1 μ M to 1 pM. Nonspecific binding was determined in the presence of 10 μ M propranolol. After incubation at 25 °C for 90 min, membranes were collected and washed via vacuum filtration (Brandel) and the bound radioactivity was quantified using a Packard Cobra gamma counter (Perkin Elmer). Equilibrium inhibition constant (K_i) values were calculated from nonlinear regression analysis (Graphpad) using the method in ref. 48.

p53 reporter assay. U2OS cells were transfected with the p53-luc reporter plasmid (Stratagene) in the presence of serum, using FuGENE6 transfection reagent (Roche). Three hours after the transfection, media were changed to serum-free media containing 100 μ M ascorbic acid and appropriate concentrations of isoproterenol. Cells were incubated for 24 h and lysed in $\times 1$ passive lysis buffer (PLB, Promega). The firefly luciferase reporter was analysed with addition of luciferase assay reagent II (Promega).

Chromatin immunoprecipitation assay (ChIP) and Re-ChIP. ChIP was performed as previously described⁴⁹. In brief, both wild-type and *Arrb1*^{-/-} MEFs were incubated with 50 μ M etoposide for 20 h. After incubation, cells were treated with 2 mM disuccinimidyl glutarate (Pierce) to crosslink protein complexes, then treated with formaldehyde to link protein to DNA covalently. Cells were lysed and the nucleoprotein complexes were sonicated. DNA-protein complexes enriched by the initial immunoprecipitation with anti-ARRB1 (K-16) antibody were eluted from beads with elution buffer (1% SDS, 0.1 M NaHCO₃), and further immunoprecipitated with anti-p53 (FL-393) antibody for Re-ChIP. The retrieved complexes were then analysed by PCR amplification of p53-binding elements in the *p21* promoter.

Statistics. Unless otherwise noted, *P* values were calculated with Student's *t*-test (two-tailed). Analysis of variance was performed with Prism (GraphPad).

1. Attramadal, H. *et al.* β -arrestin2, a novel member of the arrestin/ β -arrestin gene family. *J. Biol. Chem.* **267**, 17882–17890 (1992).
2. Xiao, K., Shenoy, S. K., Nobles, K. & Lefkowitz, R. J. Activation-dependent conformational changes in β -arrestin 2. *J. Biol. Chem.* **279**, 55744–55753 (2004).
3. Koldobskiy, M. A. *et al.* p53-mediated apoptosis requires inositol hexakisphosphate kinase-2. *Proc. Natl Acad. Sci. USA* **107**, 20947–20951 (2010).
4. Hara, M. R. *et al.* S-nitrosylated GAPDH initiates apoptotic cell death by nuclear translocation following Siah1 binding. *Nature Cell Biol.* **7**, 665–674 (2005).
5. Kohout, T. A., Lin, F. S., Perry, S. J., Conner, D. A. & Lefkowitz, R. J. β -Arrestin 1 and 2 differentially regulate heptahelical receptor signaling and trafficking. *Proc. Natl Acad. Sci. USA* **98**, 1601–1606 (2001).
6. Todaro, G. J. & Green, H. Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *J. Cell Biol.* **17**, 299–313 (1963).
7. Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).
8. Yu, J. Y., DeRuiter, S. L. & Turner, D. L. RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells. *Proc. Natl Acad. Sci. USA* **99**, 6047–6052 (2002).
9. Seshadri, T. & Campisi, J. Repression of c-fos transcription and an altered genetic program in senescent human fibroblasts. *Science* **247**, 205–209 (1990).
10. Conner, D. A. *et al.* β -Arrestin1 knockout mice appear normal but demonstrate altered cardiac responses to β -adrenergic stimulation. *Circ. Res.* **81**, 1021–1026 (1997).
11. Chruscinski, A. J. *et al.* Targeted disruption of the β_2 adrenergic receptor gene. *J. Biol. Chem.* **274**, 16694–16700 (1999).
12. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) method. *Methods* **25**, 402–408 (2001).
13. Kovacs, J. J. *et al.* β -arrestin-mediated localization of smoothened to the primary cilium. *Science* **320**, 1777–1781 (2008).
14. Lista, F., Bertness, V., Guidos, C. J., Danska, J. S. & Kirsch, I. R. The absolute number of trans-rearrangements between the *TCRG* and *TCRB* loci is predictive of lymphoma risk: a severe combined immune deficiency (SCID) murine model. *Cancer Res.* **57**, 4408–4413 (1997).
15. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
16. Hausdorff, W. P., Hnatowich, M., O'Dowd, B. F., Caron, M. G. & Lefkowitz, R. J. A mutation of the β_2 -adrenergic receptor impairs agonist activation of adenylyl cyclase without affecting high affinity agonist binding. Distinct molecular determinants of the receptor are involved in physical coupling to and functional activation of Gs. *J. Biol. Chem.* **265**, 1388–1393 (1990).
17. Baker, J. G. The selectivity of β -adrenoceptor antagonists at the human β_1 , β_2 and β_3 adrenoceptors. *Br. J. Pharmacol.* **144**, 317–322 (2005).
18. Cheng, Y. & Prusoff, W. H. Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099–3108 (1973).
19. Nowak, D. E., Tian, B. & Brasier, A. R. Two-step cross-linking method for identification of NF- κ B gene network by chromatin immunoprecipitation. *Biotechniques* **39**, 715–725 (2005).

In vitro centromere and kinetochore assembly on defined chromatin templates

Annika Guse¹, Christopher W. Carroll¹, Ben Moree¹, Colin J. Fuller¹ & Aaron F. Straight¹

During cell division, chromosomes are segregated to nascent daughter cells by attaching to the microtubules of the mitotic spindle through the kinetochore. Kinetochore are assembled on a specialized chromatin domain called the centromere, which is characterized by the replacement of nucleosomal histone H3 with the histone H3 variant centromere protein A (CENP-A). CENP-A is essential for centromere and kinetochore formation in all eukaryotes but it is unknown how CENP-A chromatin directs centromere and kinetochore assembly¹. Here we generate synthetic CENP-A chromatin that recapitulates essential steps of centromere and kinetochore assembly *in vitro*. We show that reconstituted CENP-A chromatin when added to cell-free extracts is sufficient for the assembly of centromere and kinetochore proteins, microtubule binding and stabilization, and mitotic checkpoint function. Using chromatin assembled from histone H3/CENP-A chimaeras, we demonstrate that the conserved carboxy terminus of CENP-A is necessary and sufficient for centromere and kinetochore protein recruitment and function but that the CENP-A targeting domain—required for new CENP-A histone assembly²—is not. These data show that two of the primary requirements for accurate chromosome segregation, the assembly of the kinetochore and the propagation of CENP-A chromatin, are specified by different elements in the CENP-A histone. Our unique cell-free system enables complete control and manipulation of the chromatin substrate and thus presents a powerful tool to study centromere and kinetochore assembly.

Metazoan centromeres are specified epigenetically by the presence of CENP-A nucleosomes³. Structural differences between CENP-A and histone H3 nucleosomes^{2,4} and/or specific protein recognition elements in CENP-A seem to provide the information that specifies centromere identity and directs kinetochore assembly in a DNA-sequence-independent manner^{5–10}. Moreover, many metazoan centromeres are complex in their organization, with interspersed blocks of CENP-A nucleosomes and histone H3 nucleosomes assembled on long arrays of repetitive DNA^{11–13}. The difficulty in purifying and manipulating complex centromeres has limited our understanding of how centromeric chromatin promotes centromere and kinetochore formation and chromosome segregation.

To mimic the arrays of CENP-A nucleosomes present in complex vertebrate centromeres, we reconstituted human CENP-A chromatin from recombinant components (Fig. 1a). We generated saturated chromatin arrays by salt dialysis of purified histone proteins H2A, H2B, H4 and either CENP-A or H3 with a biotinylated DNA template containing 19 repeats of a 147 bp high-affinity nucleosome positioning sequence (19X601) (Supplementary Fig. 1a, b)^{14,15}. We bound the biotinylated arrays to streptavidin-coated magnetic beads, thereby immobilizing the arrays so that they can be easily added to and recovered from cell extracts (Fig. 1a and Supplementary Fig. 1c–e).

We recently demonstrated that the essential centromere protein CENP-C directly recognizes the C terminus of CENP-A in mononucleosomes but not in isolated CENP-A₂/H4₂ tetramers⁵ (our unpublished observations). Therefore, we tested *in vitro* translated human and *Xenopus laevis* CENP-C for binding to reconstituted H3 and CENP-A

chromatin. Human and *Xenopus* CENP-A are >50% identical (Supplementary Fig. 2a) and we find that both human and *Xenopus* CENP-C bind specifically to human CENP-A chromatin arrays *in vitro*, when compared to H3 chromatin arrays (Supplementary Fig. 2b).

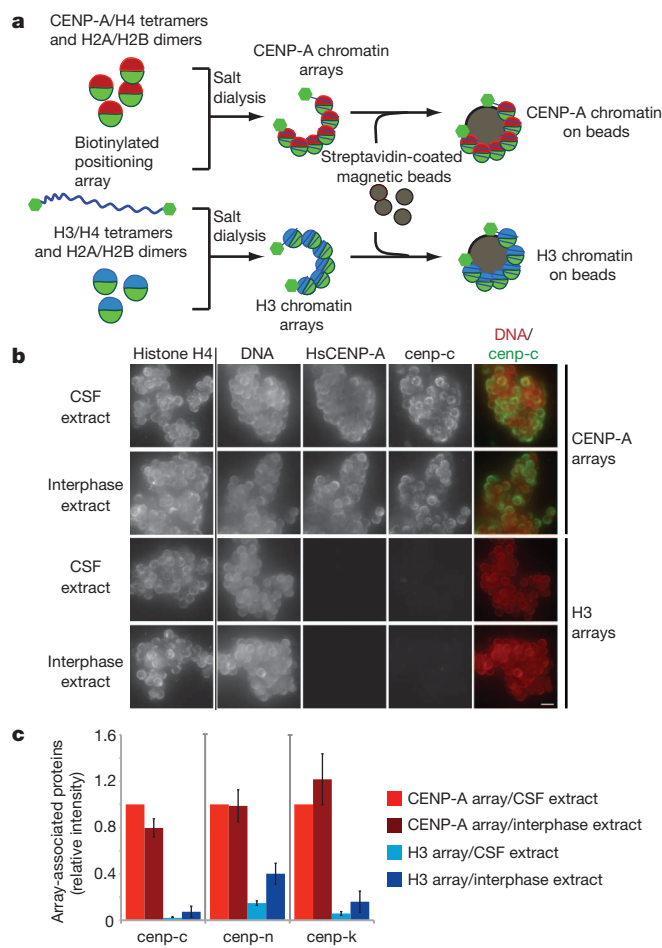


Figure 1 | Reconstituted CENP-A chromatin supports centromere assembly in *Xenopus* egg extracts. **a**, A schematic showing the reconstitution of CENP-A and H3 chromatin arrays and the attachment of the chromatin to magnetic beads via biotin end-labelled DNA. **b**, Representative images comparing cenp-c binding to human CENP-A (HsCENP-A) and H3 chromatin arrays in CSF and interphase *Xenopus* extract. The left column shows the separate histone H4 staining used for normalization of the quantification, followed by staining for DNA, human CENP-A and cenp-c. A merge image of the DNA (red) and cenp-c (green) channels is shown in the right column. Scale bar, 5 μm. **c**, Quantification of the array-associated centromeric proteins cenp-c, cenp-n and cenp-k in CSF and interphase extracts, normalized to histone H4 levels. The levels are rescaled so that CENP-A arrays in CSF are set at 1. Error bars represent the standard error of the mean (s.e.m.), $n = 3$ ($P < 0.05$ between CENP-A and H3 chromatin arrays for cenp-c, cenp-n and cenp-k).

¹Department of Biochemistry, Stanford Medical School, Beckman 409A, Stanford, California 94305-5307, USA.

Xenopus egg extract is a widely used cell-free system to study chromosome segregation¹⁶. Egg extracts are arrested in metaphase II of meiosis by the activity of cytosolic factor (CSF) and the cell-cycle state of the extract can be transitioned into interphase by adding calcium. We developed a quantitative immunofluorescence assay to determine whether centromere proteins bound to CENP-A chromatin arrays when arrays were added to *Xenopus* egg extracts. CENP-N and CENP-K are centromere proteins that are required for proper centromere and kinetochore assembly in somatic cells, and we have previously shown that CENP-N, similar to CENP-C, directly binds to the CENP-A nucleosome⁶. We found that cenp-c, cenp-n and cenp-k specifically associated with CENP-A arrays independent of the cell-cycle stage of the extract (Fig. 1b, c and Supplementary Fig. 2c–f). The centromere protein cenp-t that binds to either H3 nucleosomes or DNA at centromeres did not selectively bind CENP-A chromatin arrays (Supplementary Fig. 3a, b)¹⁷. Similarly, the inner centromere protein incenp and polo-like kinase 1 (plk1) associated with both types of chromatin arrays (Supplementary Fig. 3c). *Xenopus* incenp is targeted to chromatin through phosphorylation of both H2A and H3 and thus may have affinity for both CENP-A and H3 chromatin^{18–20} and plk1 associates with chromatin in *Xenopus* egg extract independent of the kinetochore²¹. Furthermore, reconstituted chromatin segments are unlikely to generate paired sister chromatids with inner centromeres because naked DNA and linear DNA replicates inefficiently in these egg extracts²². The specific recruitment of the centromere proteins cenp-c, cenp-n and cenp-k, however, indicates that reconstituted CENP-A chromatin arrays can support essential steps in the centromere assembly process *in vitro*.

Functional kinetochores assemble on sperm chromatin in metaphase *Xenopus* egg extract. At high sperm concentration, microtubule depolymerization causes mitotic checkpoint activation, resulting in the increased association of checkpoint proteins with kinetochores and cell-cycle arrest²³. We tested whether reconstituted CENP-A chromatin arrays support kinetochore assembly and checkpoint protein binding after microtubule depolymerization. We added CENP-A or H3 arrays to CSF-arrested egg extracts and then cycled the extracts through interphase and back into mitosis, in the presence or absence of nocodazole, as outlined in Fig. 2a and demonstrated in Supplementary Fig. 4a. The constitutive centromere protein cenp-c and the microtubule-binding kinetochore protein ndc80 bound to CENP-A arrays in the presence or absence of nocodazole (Fig. 2b, c and Supplementary Fig. 4b). The spindle assembly checkpoint proteins cenp-e, mad2, rod (also known as kntc1) and zw10 associated with CENP-A chromatin at intermediate levels in the absence of nocodazole but upon microtubule depolymerization their binding increased 2–4 fold (Fig. 2b). Western blot analysis showed that cenp-c and ndc80 are precipitated with CENP-A arrays independent of microtubule depolymerization. *Xenopus* zw10 and rod are enriched on CENP-A arrays upon nocodazole treatment in metaphase, regardless of whether the extract has been cycled through interphase (Fig. 2c). These results indicate that CENP-A chromatin arrays respond to microtubule depolymerization by recruiting mitotic checkpoint proteins (Fig. 2b, c and Supplementary Fig. 4b).

Microtubule binding is a hallmark of kinetochore function and decondensed sperm chromatin efficiently supports spindle formation in egg extracts (Fig. 3a, left)²⁴. However, chromatin assembled on naked DNA induces spindle formation in *Xenopus* egg extracts independent of kinetochores²⁵. When we added CENP-A and H3 chromatin beads into mitotic egg extract we observed microtubule polymerization around the majority of CENP-A arrays but only around a subset of H3 arrays (Fig. 3a, left). We quantified the amount of microtubule polymer associated with each type of array and found significantly more microtubules associated with CENP-A chromatin beads (Fig. 3b and Supplementary Fig. 5a). This indicates that CENP-A chromatin preferentially stabilizes microtubules or promotes their polymerization. We observed heterogeneous microtubule structures around the CENP-A chromatin beads ranging from bipolar spindles to stabilized

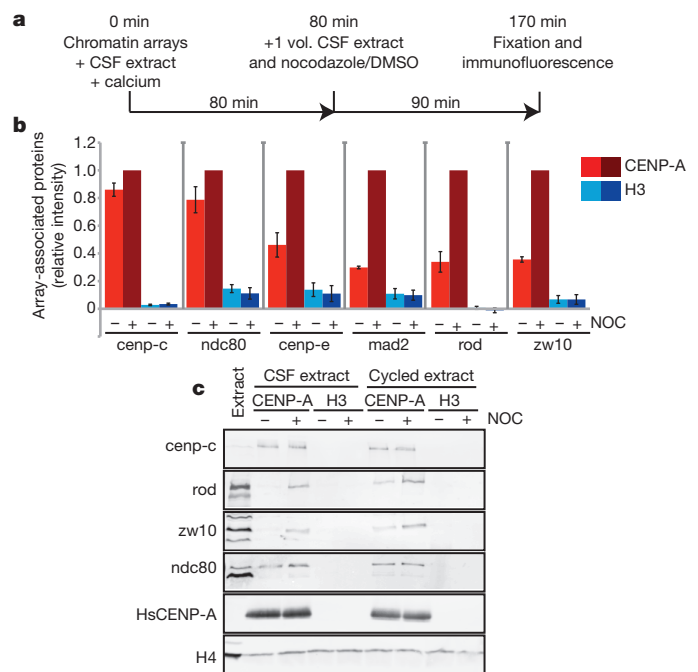


Figure 2 | CENP-A chromatin specifically recruits kinetochore proteins as a response to a mimic of kinetochore detachment from microtubules. **a**, A schematic showing the experimental procedure. **b**, Quantification of immunofluorescence analysis of cenp-c, ndc80, cenp-e, mad2, rod or zw10 recruitment to chromatin arrays with (+) and without (–) nocodazole (NOC). The levels are rescaled so that CENP-A arrays with nocodazole are set at 1. Error bars represent s.e.m., $n = 3$ ($P < 0.05$ between (–) and (+) nocodazole for cenp-e, mad2, rod and zw10 binding to CENP-A chromatin arrays). **c**, Western blot analysis of cenp-c, ndc80, rod and zw10 recruitment to CENP-A (HsCENP-A) and H3 chromatin arrays with and without nocodazole in CSF and cycled egg extracts. H4 levels are shown as a loading control.

microtubules or microtubule bundles (Fig. 3a and Supplementary Fig. 5a, b). A second property of functional kinetochores is that kinetochore-associated microtubule bundles (k-fibres) are stable to cold treatment, which depolymerizes non-kinetochore microtubules. We asked whether kinetochores assembled on CENP-A chromatin could stabilize microtubules to cold shock by incubating the microtubule assembly reactions for 10 min at 4 °C. We found that kinetochores assembled on CENP-A chromatin arrays stabilized microtubules to cold shock similar to kinetochores assembled on native sperm chromatin whereas H3 chromatin arrays did not (Fig. 3a, c and Supplementary Fig. 5c). When we completely depolymerized microtubules with nocodazole we observed mad2 recruitment to native sperm centromeres and CENP-A chromatin beads but not H3 chromatin beads (Fig. 3a, c and Supplementary Fig. 5c). These results indicate that CENP-A chromatin arrays, similar to native sperm chromatin, assemble functional kinetochores that promote microtubule binding, k-fibre stabilization and spindle checkpoint function (Fig. 3a).

In cells, unattached kinetochores activate the mitotic checkpoint and delay mitotic exit until all chromosomes are properly attached and aligned^{26,27}. We tested whether kinetochores assembled on CENP-A chromatin arrays could generate a mitotic checkpoint response to microtubule depolymerization and delay the cell cycle. We mixed CENP-A and H3 chromatin with CSF extracts, cycled the reactions through interphase and then cycled them back into mitosis in the presence or absence of nocodazole (Fig. 2a). We then released the extract from mitosis into interphase a second time and monitored the kinetics of this transition by measuring the mitosis-specific phosphorylation of wee1 (phospho-wee1) (Fig. 3d). On release from mitosis, phospho-wee1 levels rapidly declined and were undetectable after 30 min in control extracts containing CENP-A chromatin or H3 chromatin, as well as in extracts containing H3 chromatin in the presence of

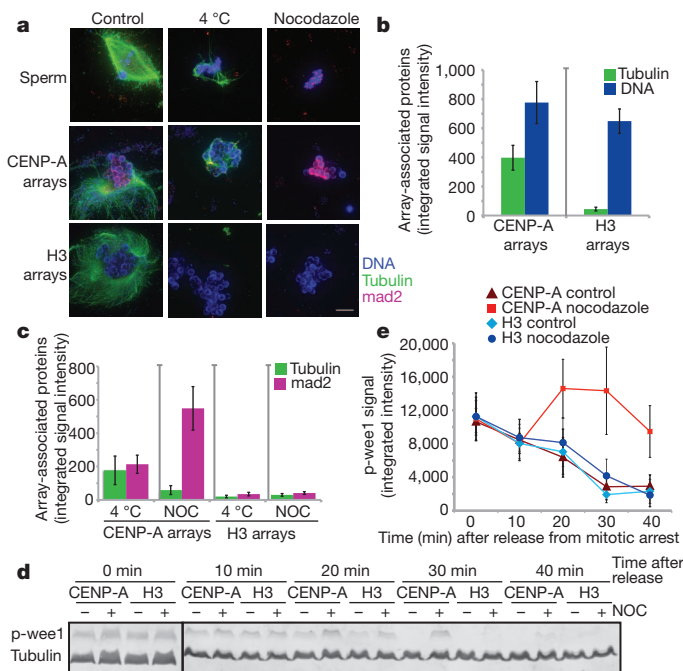


Figure 3 | Kinetochore assembly on reconstituted CENP-A chromatin bind microtubules and generate a mitotic checkpoint signal.

a, Representative images of microtubule polymerization induced by sperm or reconstituted CENP-A and H3 chromatin. Microtubules (green) and mad2 (magenta) levels are shown. Scale bar, 10 μ m. **b**, Quantification of tubulin and DNA associated with CENP-A and H3 chromatin beads. Error bars represent s.e.m., $n = 5$. **c**, Quantification of tubulin and mad2 levels associated with CENP-A and H3 chromatin beads after cold shock (4 °C) and nocodazole (NOC) treatment. Error bars represent s.e.m., $n = 5$. **d**, Western blot showing phospho-wee1 (p-wee1) levels as an indicator of the cell-cycle stage and tubulin levels as a loading control. Samples from different time points after release from mitotic arrest are shown for CENP-A and H3 chromatin arrays, each incubated with nocodazole (+) or with DMSO (–) as a control. **e**, Quantification of four independent experiments showing the phospho-wee1 signal intensity (p-wee1 signal) over time (min). Error bars represent s.e.m., $n = 4$.

nocodazole (Fig. 3d, e). In extracts containing CENP-A chromatin and nocodazole, the phospho-wee1 signal increased until 20 min after calcium addition and then declined until 40 min after calcium addition to a level only slightly lower than that before release (Fig. 3d, e). In the presence of CENP-A chromatin and nocodazole, cyclin B levels rapidly declined but then stabilized, similar to the response observed for native sperm chromatin²³. However, cyclin B was not stabilized in the presence of H3 chromatin and nocodazole (Supplementary Fig. 5d, e). We estimate that the number of CENP-A nucleosomes we are adding to the egg extract exceeds the CENP-A nucleosome concentration required to activate the checkpoint using sperm nuclei²³. The lower efficiency of reconstituted arrays for checkpoint signalling may be due to the comparatively short length of our reconstituted CENP-A chromatin to native CENP-A chromatin or the lack of replicated sister chromatids and inner centromeres important for tension-dependent checkpoint activation. Despite these differences, our synthetic CENP-A chromatin supports a mitotic checkpoint response that mimics the response of native kinetochores to microtubule depolymerization.

The reconstituted chromatin system we have developed provides a distinct experimental advantage over native metazoan centromeric chromatin because the chromatin template can be easily manipulated to dissect the roles of histone proteins in centromere function. A central question in centromere function is how CENP-A chromatin directs the assembly of the centromere and kinetochore. CENP-N recognizes the CATD region of the CENP-A nucleosome while CENP-C binds the C-terminal tail of CENP-A^{5,6}. However, the relative importance of these two recognition mechanisms in centromere and kinetochore assembly is incompletely understood.

We generated chromatin arrays containing chimaeric CENP-A/H3 proteins to ask how the CENP-A CATD domain and the CENP-A C terminus influence centromere and kinetochore assembly (Fig. 4a). We characterized the level of histone exchange and/or loss from the arrays during incubation in extracts and found that the majority of recombinant human CENP-A nucleosomes were stable during the incubation, indicating low exchange and/or loss rates (Supplementary Fig. 6a, b). We detected a low level of phosphorylated histone H3 on CENP-A chromatin arrays in CSF extract ($11.7\% \pm 7\%$ compared to H3 arrays) and in extract that had been cycled through interphase and back into mitosis ($22\% \pm 13\%$ compared to H3 arrays) (Supplementary Fig. 6c, d). The chimaeric arrays containing CENP-A with the histone H3 tail (CENP-A + H3C) exhibited similar levels of exchange (Supplementary Fig. 6c, d). The *Xenopus* cenp-a present in the extract did not appreciably exchange onto any of the arrays (detection limit ~ 5 –10% exchange) (Supplementary Fig. 6c). The absence of gross rearrangements or bulk histone exchange suggests that chromatin arrays can be used to dissect how individual domains of CENP-A influence kinetochore assembly.

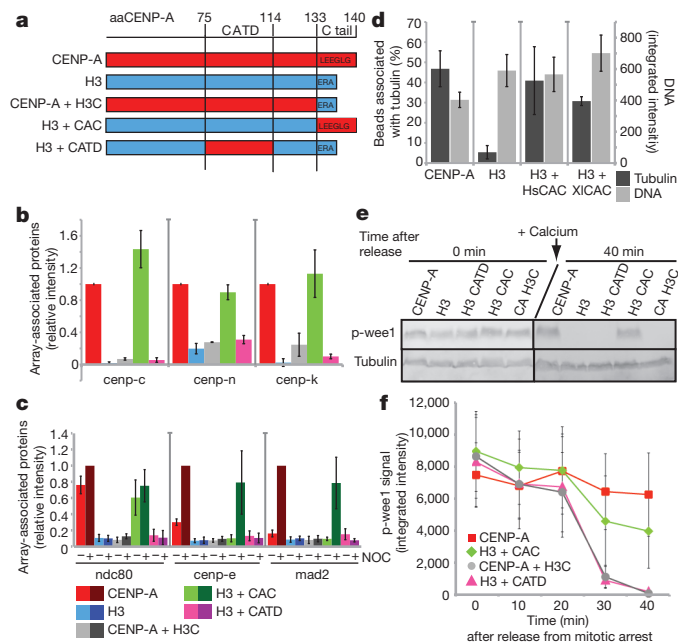


Figure 4 | The CENP-A C terminus is required for centromere and kinetochore assembly in *Xenopus* egg extract.

a, A schematic showing the different CENP-A/H3 chimaeras used in this study. The numbers at the top represent the amino acid (aa) within human CENP-A. **b**, Quantification of immunofluorescence analysis of cenp-c, cenp-k and cenp-n recruitment to wild-type and chimaeric arrays. The relative amounts of each centromere protein bound to the arrays are shown relative to CENP-A arrays set to 1. Error bars represent s.e.m., $n = 3$ ($P \leq 0.05$ for all proteins binding to CENP-A arrays compared to chimaeric arrays except for the H3 arrays containing the CENP-A C terminus). **c**, Quantification of immunofluorescence analysis of ndc80, cenp-e, mad2 recruitment to chimaeric chromatin arrays with (+) and without (–) nocodazole (NOC). Values are displayed relative to CENP-A arrays in the presence of nocodazole set to 1. Error bars represent s.e.m., $n = 4$. The efficiencies of recruitment of kinetochore proteins to CENP-A and H3 + CAC arrays in nocodazole were not statistically distinguishable ($P \geq 0.26$ for ndc80, cenp-e and mad2). **d**, Quantification of microtubule binding to CENP-A, H3, H3 + human CAC (HsCAC) and H3 + *Xenopus* CAC (XICAC) chromatin arrays represented as percentage of beads associated with tubulin levels above threshold (dark grey bars, left y-axis). Average DNA levels on chromatin beads are shown representing the levels of chromatin arrays bound to beads (light grey bars, right y-axis). Error bars represent s.e.m., $n = 4$ for CENP-A and H3 arrays, $n = 5$ for H3 + human CAC arrays and $n = 2$ for H3 + *Xenopus* CAC arrays. **e**, Western blot analysis shows phospho-wee1 (p-wee1) levels as an indicator of the cell-cycle stage at 0 min and 40 min after mitotic exit. Tubulin levels are shown as a loading control. **f**, Quantification of the phospho-wee1 signal intensity over time. Error bars represent s.e.m., $n = 5$.

Using our *in vitro* centromere and kinetochore assembly assay, we found that cenp-c bound with equal efficiency to chromatin arrays assembled with either wild-type CENP-A or with chimaeras of histone H3 with the CENP-A C-terminal six amino acids (H3 + CAC) but not CENP-A + H3C (Fig. 4b and Supplementary Fig. 7a, left). This demonstrates that the CENP-A C terminus is necessary and sufficient for recruiting cenp-c to CENP-A chromatin arrays in egg extracts, as it is for CENP-A mononucleosome binding *in vitro*⁵.

Xenopus cenp-k depends on cenp-c for its association with sperm centromeres²⁸ and cenp-k also associated with the wild-type and H3 + CAC arrays (Fig. 4b and Supplementary Fig. 7a). Surprisingly, we found that H3 + CAC arrays recruited cenp-n as efficiently as wild-type CENP-A arrays, even though these arrays lack the CATD recognition element for CENP-N⁶. *Xenopus* cenp-n binding to either CENP-A + H3C or H3 + CATD arrays was no better than its binding to H3 chromatin arrays, indicating that the CENP-A C terminus is required for cenp-n association with CENP-A chromatin in *Xenopus* egg extract (Fig. 4b and Supplementary Fig. 7a). The lack of *Xenopus* cenp-n binding to H3 + CATD and CENP-A + H3C chromatin arrays is not due to species differences because *Xenopus* cenp-n binds human CENP-A mononucleosomes *in vitro* in the absence of CENP-C (Supplementary Fig. 7b). The association of cenp-n and cenp-k with chromatin arrays was dependent on cenp-c, as cenp-c depletion from the extract (Supplementary Fig. 8a) reduced the binding to background levels (Supplementary Fig. 8b, c). This was not due to depletion of cenp-n or cenp-k by cenp-c, as we have previously shown that complementation of cenp-c-depleted extracts restores cenp-k binding and CENP-K is known to depend on CENP-N for its centromere localization^{6,28,29,30}. The dependence of CENP-N on CENP-C for its localization to CENP-A arrays may reflect a role for CENP-C in altering the geometry of centromeric chromatin to promote access of CENP-N to CENP-A nucleosomes, or it may reflect the assembly of CENP-N into the larger CCAN complex recruited to the centromere via CENP-C. Our results demonstrate that cenp-c recognition of the CENP-A C terminus is necessary and sufficient for cenp-n and cenp-k association with chromatin arrays in *Xenopus* egg extract.

We analysed the chromatin requirements for mitotic kinetochore formation using the experimental strategy illustrated in Fig. 2a. The kinetochore proteins ndc80, cenp-e, mad2, rod and zw10 are efficiently recruited to wild-type and H3 + CAC chromatin arrays, but not to CENP-A + H3C or H3 + CATD chromatin arrays (Fig. 4c, Supplementary Fig. 9a and Supplementary Fig. 10a). Similar to wild-type CENP-A chromatin, only the checkpoint proteins cenp-e, mad2, zw10 and rod increased in their association with H3 + CAC after microtubule depolymerization (Fig. 4c, Supplementary Fig. 9a and Supplementary Fig. 10a). As with wild-type CENP-A arrays, the H3 + CAC arrays showed increased associated microtubule polymer indicating that the C terminus of CENP-A directs the formation of microtubule binding or stabilization activity (Fig. 4d). Human and *Xenopus* CENP-A differ by two amino acids in their C-terminal tail (Supplementary Fig. 2a) and chimaeric nucleosome arrays containing the *Xenopus* C-terminal tail of cenp-a fused to H3 (H3 + *Xenopus* CAC) were equally efficient in cenp-c recruitment and microtubule binding as human H3 + CAC arrays (Fig. 4d and Supplementary Fig. 10b); indicating that the mode of interaction between CENP-C and CENP-A is conserved.

We assayed the ability of chimaeric nucleosome arrays to promote mitotic checkpoint arrest after microtubule depolymerization and found that H3 + CATD and CENP-A + H3C did not delay the exit from mitosis but that H3 + CAC did (Fig. 4e, f). The delay of mitotic exit caused by H3 + CAC arrays was less effective than that of CENP-A chromatin arrays, indicating that regions of CENP-A in addition to the C terminus increase the effectiveness of checkpoint signalling, possibly by stabilizing CCAN and kinetochore protein interactions with chromatin (Fig. 4e, f). Taken together, our data demonstrate that the primary chromatin determinant for functional centromere and

kinetochore assembly is the C terminus of CENP-A and its recognition by CENP-C.

Here we have shown that reconstituted CENP-A chromatin, in the absence of native centromeric DNA, is necessary and sufficient for centromere and kinetochore assembly. Our data imply that short domains of CENP-A chromatin are sufficient for assembling core components of the centromere and kinetochore in the absence of higher-order organization of centromeric chromatin and interspersed domains of H3 chromatin.

Using our *in vitro* system, we have directly assessed how domains of CENP-A participate in centromere and kinetochore assembly, even when the mutations we analyse would be expected to be lethal *in vivo*. We find that the CENP-A C terminus is both necessary and sufficient for the recruitment of centromere and kinetochore proteins, for microtubule binding and for a checkpoint response to microtubule depolymerization. We suggest that CENP-A performs two functions that can be separated molecularly: (1) the CENP-A CATD provides a recognition mechanism for targeting of CENP-A to centromeres to maintain centromeric chromatin^{2,6–8}; and (2) the CENP-A C-terminal tail domain recruits the conserved centromere protein CENP-C to promote centromere and kinetochore assembly⁵. We envision the use of more complex chromatin templates to understand the importance of higher-order chromatin organization and regulatory modifications in centromere assembly and function.

METHODS SUMMARY

Histone proteins and chimaeras were purified as described previously^{5,6,15} and assembled onto a biotin end-labelled tandem array of 19 high-affinity nucleosome positioning sequences (19X601) by salt dialysis¹⁴. Chromatin arrays were bound to streptavidin-coated magnetic Dynabeads (Invitrogen). *X. laevis* extracts were prepared as previously described¹⁶ and centromere protein binding to chromatin arrays was performed in freshly prepared CSF egg extract for 1 h with or without calcium addition. Arrays were fixed in formaldehyde and stained for centromere proteins by indirect immunofluorescence. Kinetochore and checkpoint protein assembly was assayed by adding arrays to extracts released into interphase with calcium for 80 min followed by re-addition of CSF extract in the presence or absence of nocodazole (10 $\mu\text{g ml}^{-1}$) for another 90 min. To analyse microtubule binding, chromatin arrays were incubated in CSF for 90 min. Reactions were sedimented through a glycerol cushion onto a coverslip followed by tubulin immunofluorescence. Chromatin-array-dependent inhibition of mitotic exit was assayed as described for kinetochore protein binding, but calcium was added a second time to release extracts into interphase. The cell-cycle state was monitored by western blotting using anti-phospho-wee1 antibody, provided by J. E. Ferrell.

Images were collected as 13 axial planes at 2 μm intervals on a Nikon Eclipse-80i microscope using a $\times 60$, 1.4 NA PlanApo oil lens and a CoolSnapHQ CCD camera (Photometrics) with MetaMorph software (MDS Analytical Technologies). Axial stacks were maximum intensity projected and quantified using custom software. For normalization of each experiment, a separate histone H4 staining was performed to quantify the exact array coupling efficiency.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 November 2010; accepted 19 July 2011.

Published online 28 August 2011.

- Cheeseman, I. M. & Desai, A. Molecular architecture of the kinetochore-microtubule interface. *Nature Rev. Mol. Cell Biol.* **9**, 33–46 (2008).
- Black, B. E. *et al.* Structural determinants for generating centromeric chromatin. *Nature* **430**, 578–582 (2004).
- Black, B. E. & Bassett, E. A. The histone variant CENP-A and centromere specification. *Curr. Opin. Cell Biol.* **20**, 91–100 (2008).
- Sekulic, N., Bassett, E. A., Rogers, D. J. & Black, B. E. The structure of (CENP-A-H4)₂ reveals physical features that mark centromeres. *Nature* **467**, 347–351 (2010).
- Carroll, C. W., Milks, K. J. & Straight, A. F. Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J. Cell Biol.* **189**, 1143–1155 (2010).
- Carroll, C. W., Silva, M. C. C., Godek, K. M., Jansen, L. E. T. & Straight, A. F. Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. *Nature Cell Biol.* **11**, 896–902 (2009).
- Dunleavy, E. M. *et al.* HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* **137**, 485–497 (2009).
- Foltz, D. R. *et al.* Centromere-specific assembly of CENP-A nucleosomes is mediated by HJURP. *Cell* **137**, 472–484 (2009).

9. Hu, H. *et al.* Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. *Genes Dev.* **25**, 901–906 (2011).
10. Tachiwana, H. *et al.* Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature*. doi:10.1038/nature10258 (10 July, 2011).
11. Blower, M. D., Sullivan, B. A. & Karpen, G. H. Conserved organization of centromeric chromatin in flies and humans. *Dev. Cell* **2**, 319–330 (2002).
12. Ribeiro, S. *et al.* A super-resolution map of the vertebrate kinetochore. *Proc. Natl Acad. Sci. USA* **107**, 10484–10489 (2010).
13. Zinkowski, R. P., Meyne, J. & Brinkley, B. R. The centromere-kinetochore complex: a repeat subunit model. *J. Cell Biol.* **113**, 1091–1110 (1991).
14. Huynh, V., Robinson, P. & Rhodes, D. A method for the *in vitro* reconstitution of a defined “30 nm” chromatin fibre containing stoichiometric amounts of the linker histone. *J. Mol. Biol.* **345**, 957–968 (2005).
15. Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M. & Richmond, T. J. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* **272**, 301–311 (1997).
16. Desai, A., Murray, A., Mitchison, T. J. & Walczak, C. E. The use of *Xenopus* egg extracts to study mitotic spindle assembly and function *in vitro*. *Methods Cell Biol.* **61**, 385–412 (1999).
17. Hori, T. *et al.* CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell* **135**, 1039–1052 (2008).
18. Kawashima, S. A., Yamagishi, Y., Honda, T., Ishiguro, K.-i. & Watanabe, Y. Phosphorylation of H2A by Bub1 prevents chromosomal instability through localizing shugoshin. *Science* **327**, 172–177 (2010).
19. Kelly, A. E. *et al.* Survivin reads phosphorylated histone H3 threonine 3 to activate the mitotic kinase Aurora B. *Science* **330**, 235–239 (2010).
20. Wang, F. *et al.* Histone H3 Thr-3 phosphorylation by Haspin positions Aurora B at centromeres in mitosis. *Science* **330**, 231–235 (2010).
21. Budde, P. P., Kumagai, A., Dunphy, W. G. & Heald, R. Regulation of Op18 during spindle assembly in *Xenopus* egg extracts. *J. Cell Biol.* **153**, 149–158 (2001).
22. Blow, J. J. & Laskey, R. A. Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of *Xenopus* eggs. *Cell* **47**, 577–587 (1986).
23. Minshull, J., Sun, H., Tonks, N. K. & Murray, A. W. A MAP kinase-dependent spindle assembly checkpoint in *Xenopus* egg extracts. *Cell* **79**, 475–486 (1994).
24. Sawin, K. E. & Mitchison, T. J. Mitotic spindle assembly by two different pathways *in vitro*. *J. Cell Biol.* **112**, 925–940 (1991).
25. Heald, R. *et al.* Self-organization of microtubules into bipolar spindles around artificial chromosomes in *Xenopus* egg extracts. *Nature* **382**, 420–425 (1996).
26. Nicklas, R. B., Ward, S. C. & Gorbsky, G. J. Kinetochore chemistry is sensitive to tension and may link mitotic forces to a cell cycle checkpoint. *J. Cell Biol.* **130**, 929–939 (1995).
27. Rieder, C. L., Cole, R. W., Khodjakov, A. & Sluder, G. The checkpoint delaying anaphase in response to chromosome monoorientation is mediated by an inhibitory signal produced by unattached kinetochores. *J. Cell Biol.* **130**, 941–948 (1995).
28. Milks, K. J., Moree, B. & Straight, A. F. Dissection of CENP-C-directed centromere and kinetochore assembly. *Mol. Biol. Cell* **20**, 4246–4255 (2009).
29. Foltz, D. R. *et al.* The human CENP-A centromeric nucleosome-associated complex. *Nature Cell Biol.* **8**, 458–469 (2006).
30. McClelland, S. E. *et al.* The CENP-A NAC/CAD kinetochore complex controls chromosome congression and spindle bipolarity. *EMBO J.* **26**, 5033–5047 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors would like to thank A.F.S. laboratory members for support and comments, J. E. Ferrell, A. Murray, R.-H. Chen, G. Kops and P. T. Stukenberg for providing antibodies. D. Rhodes, P. Robinson, K. Luger, J. Hansen, G. Narlikar and J. Yang for providing reagents and advice. A.G. was supported by a postdoctoral fellowship from the German Research Foundation (DFG). C.W.C. was supported by a postdoctoral fellowship from the Helen Hay Whitney Foundation and the American Heart Association (AHA). B.M. was supported by T32GM007276. C.J.F. was supported by a Stanford Graduate Fellowship and this work was supported by National Institutes of Health (NIH) R01GM074728 to A.F.S.

Author Contributions A.G. and A.F.S. designed the experiments and wrote the manuscript. A.G. performed all the experiments. C.W.C. purified the CENP-A/H3 chimaeras and assembled arrays containing chimaeric proteins, analysed *Xenopus* cenp-n binding to human CENP-A mononucleosomes and provided advice. B.M. generated *Xenopus* centromere protein antibodies and C.J.F. designed and wrote the image analysis software for quantitative analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.F.S. (astraight@stanford.edu).

METHODS

Histone expression. CENP-A/H4 and H3/H4 wild-type and chimaeric tetramers, as well as H2A and H2B dimers were expressed and purified as described previously^{5,6,15,31}. **Preparation of biotinylated array DNA.** A tandem array of 19 copies of the high-affinity nucleosome positioning sequence (19X601)^{14,32} was digested with EcoRI, XbaI, DraI and HaeII (NEB) overnight to excise the 19-nucleosome positioning sequence array and to digest the remaining backbone DNA to smaller DNA fragments. The array DNA was then purified by PEG precipitation and dialysed against 10 mM Tris-HCl pH 8.0, 0.25 mM EDTA as previously described¹⁴.

The array DNA was end labelled with biotin by end filling the EcoRI and XbaI sites using Klenow DNA polymerase for 4 h at 37 °C in a reaction containing 35 µM Biotin-14-dATP (Invitrogen), α -thio-dTTP and α -thio-dGTP (Chemcyte) and dCTP. The labelled DNA was then purified using a PCR fragment purification kit (Qiagen). The biotinylation efficiency was determined by adding FITC-streptavidin (final concentration of 10 µg ml⁻¹) to 500 ng of purified array DNA and monitoring the fraction of gel-shifted DNA after migration in a 0.7% agarose gel.

Chromatin array assembly. To assemble chromatin arrays, biotinylated DNA, CENP-A/H4 or H3/H4 tetramers and H2A/H2B dimers were mixed at a stoichiometry of 1:1:2.2 or 1:0.9:2.2, respectively, in high-salt buffer (10 mM Tris-HCl pH 7.5, 0.25 mM EDTA, 2 M NaCl) and then dialysed into low-salt buffer (10 mM Tris-HCl pH 7.5, 0.25 mM EDTA, 2.5 mM NaCl) over 60–70 h at 4 °C. Final array DNA concentration typically was 0.15 mg ml⁻¹ to 0.2 mg ml⁻¹.

To assess the efficiency of nucleosome assemblies, arrays were digested at room temperature (approximately 22 °C) overnight with Aval in a low-magnesium buffer (50 mM potassium acetate, 20 mM Tris-acetate, 0.5 mM magnesium acetate, 1 mM dithiothreitol, pH 7.9). Digested chromatin arrays were supplemented with glycerol (20% final concentration) and separated on a native 5% acrylamide gel in 0.5× Tris/Borate/EDTA buffer for 80 min at 10 mA. Gels were stained with EtBr (1 µg ml⁻¹) to visualize DNA.

Coupling of biotinylated chromatin arrays to Dynabeads. Biotinylated chromatin arrays were coupled to prewashed streptavidin-coated magnetic Dynabeads (Invitrogen) at a ratio of 10 µg DNA to 1 mg beads in 50 mM Tris-HCl pH 8.0, 75 mM NaCl, 0.25 mM EDTA, 2.5% polyvinyl alcohol (PVA) and 0.05% Triton-X-100 for 1–2 h. The beads were then equilibrated in 75 mM Tris-HCl pH 8.0, 75 mM NaCl, 0.25 mM EDTA, 0.05% Triton-X-100 and either used directly or stored at 4 °C for later use.

***X. laevis* egg extracts.** *X. laevis* CSF extracts were prepared as previously described^{16,33}. To assess the binding of centromeric proteins to chromatin arrays in CSF and interphase egg extracts, chromatin arrays were mixed with freshly prepared CSF egg extract with or without CaCl₂ (final concentration 0.6 mM) at a nucleosome concentration of ~100 nM unless stated otherwise. The reactions were incubated for 1 h at 4 °C or at 16–20 °C in a water bath, the arrays were re-isolated from extracts by exposure to a magnet and then washed three times in 1× CSF-XB buffer (10 mM HEPES pH 7.7, 2 mM MgCl₂, 0.1 mM CaCl₂, 100 mM KCl, 5 mM EGTA, 50 mM sucrose) supplemented with 0.05% Triton-X-100. Chromatin arrays were fixed in CSF-XB buffer, 0.05% Triton-X-100, 2% formaldehyde for 5 min. After fixation, chromatin arrays were washed into antibody dilution buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Triton-X-100, 2% BSA) and analysed by immunofluorescence.

Kinetochore and spindle checkpoint protein assembly were analysed by mixing chromatin arrays with CSF extract and CaCl₂ (final concentration 0.6 mM). Reactions were incubated at 16–20 °C for 80 min to allow extracts to release into interphase and mixed every 15 min. One volume of fresh CSF extract was added together with nocodazole (or DMSO) at 10 µg ml⁻¹ and samples were held at 16–20 °C for another 90 min. After 170 min total incubation time, samples for immunofluorescence analysis were washed and fixed as described above.

The cell-cycle state was verified by loading 2 µl extract of all relevant time points onto SDS-PAGE, followed by western blotting using the anti-phospho-wee1 antibody³⁴.

To assess the ability of chromatin arrays to inhibit mitotic exit, arrays were mixed with CSF extract and CaCl₂ (final concentration: 0.6 mM). The samples were incubated for 80 min to induce the release into interphase. In the next step, one volume of fresh CSF extract, supplemented with nocodazole/DMSO, was added to cycle the extract back into a mitotic arrest. After 90 min, CaCl₂ was added again to release the extract from mitotic arrest. Western blot samples were taken at all indicated time points and processed as described.

To analyse microtubule binding by CENP-A and H3 chromatin arrays, chromatin arrays were mixed with CSF extract and incubated for 90 min at 18–20 °C. During incubation samples were mixed every 15 min. Reactions were fixed for 10 min in 2.5% formaldehyde, sedimented through a glycerol cushion onto coverslips and post-fixed for 5 min in ice-cold methanol followed by immunofluorescence analysis³⁵. To assay for mad2 levels and microtubule stabilization, reactions were either supplemented with nocodazole at a final concentration of 10 µg ml⁻¹ or shifted to 4 °C for 10 min after the 90 min incubation time.

Immunodepletion. Depletion of *Xenopus* cenp-c from *Xenopus* egg extracts was performed as described previously²⁸.

Cloning and antibody generation. The *X. laevis* cenp-n cDNA clone (GenBank accession number BC084956) was purchased from American Type Culture Collection. Peptides against *Xenopus* cenp-n (acetyl-CPHKARNSFKITEKR-amide) were synthesized by Bio-Synthesis and peptide antibodies were generated as previously described³⁶.

Immunofluorescence. For immunofluorescence analysis, fixed chromatin arrays were bound to poly-L-lysine-coated acid-washed coverslips. The following primary antibodies were used for immunofluorescence staining and typically incubated at 4 °C overnight: anti-human CENP-A³⁰ was directly coupled to Alexa 647 (Molecular Probes), anti-H4 (Abcam), anti-*Xenopus* cenp-c, anti-*Xenopus* cenp-e, anti-*Xenopus* cenp-k and anti-*Xenopus* cenp-n and anti-tubulin (Dm1 α ; Sigma). Rabbit antibodies were generated against the full-length *Xenopus* polo kinase made in Sf9 cells and a GST fusion to the first 379 amino acids of *Xenopus* incenp made in *E. coli*. The anti-mad2 antibody was provided by A. Murray (Harvard University), and R.-H. Chen (Institute of Molecular Biology, Academia Sinica), the anti-*Xenopus* zw10 and anti-*Xenopus* rod antibodies were provided by G. Kops (University Medical Center Utrecht) and the anti-*Xenopus* ndc80 antibody was provided by P. Todd Stukenberg (University of Virginia). Alexa-conjugated secondary antibodies were used at 1 µg ml⁻¹ (Molecular Probes). Propidium iodide at 1 µg ml⁻¹ or Hoechst at 10 µg ml⁻¹ was used to visualize DNA.

Microscopy and analysis. Images were collected on a Nikon Eclipse 80i microscope using a ×60, 1.4 NA Plan Apo VC oil immersion lens, a Sedat Quad filter set (Chroma Technology) using MetaMorph software (MDS Analytical Technologies) and a charge-coupled device camera (CoolSnapHQ; Photometrics). Thirteen axial planes at 2 µm intervals were acquired with an MFC-2000 Z-axis drive (Applied Scientific Instrumentation). Axial stacks were maximum intensity projected and then quantified using custom software (Matlab) to identify beads in each image and to quantify the integrated intensity for each channel after background subtraction. Briefly, the propidium iodide stained (DNA) channel was used to find beads. Bead centroids were found by filtering the image using a structuring element that had a peak at a 17 pixel radial distance from the structuring element centre, corresponding to the bright ring seen around the edges of the beads. A 35 pixel diameter circle around the centroid of each bead identified was used as the region of interest for that bead. After beads were identified, regions of interest were transferred automatically to the remaining channels and the integrated signal intensity was calculated for each bead in each channel, normalized to the area of the bead region (which was uniform except in cases of partially overlapping beads), and background corrected using an average of three bead-sized regions manually chosen to be away from any beads. For each experiment, at least three images per coverslip were acquired and 20–300 beads were analysed per image. For the normalization of each experiment, a separate histone H4 staining was performed to quantify the exact coupling efficiency for each type of chromatin array and for each experiment.

Immunofluorescence microscopy images of the microtubule binding assays that were subjected to deconvolution were acquired with an Olympus IX70 microscope. The microscope was outfitted with a Deltavision Core system (Applied Precision) using an Olympus ×60 1.4NA Plan Apo lens, a Sedat Quad filter set (Semrock) and a CoolSnap HQ CCD Camera (Photometrics). The microscope was controlled via softwrx 4.1.0 software (Applied Precision) and images were deconvolved using softwrx v. 4.1.0 (Applied Precision). Microtubule quantification was performed using a modification of the same software used for centromere protein quantification.

Immunoblotting. Western blot samples were separated by SDS-PAGE and transferred onto PVDF membrane (Bio-Rad) in CAPS transfer buffer (10 mM 3-(cyclohexylamino)-1-propanesulfonic acid, pH 11.3, 0.1% SDS and 20% methanol). The following primary antibodies were typically incubated overnight at 4 °C: anti-*Xenopus* cenp-c²⁸, anti-tubulin (Dm1 α , Sigma), anti-H4 (Abcam), anti-phospho H3 (Ser10) (Millipore), anti-phospho-wee1. The anti-phospho-wee1 antibody was provided by J. E. Ferrell (Stanford University)³⁴. For additional primary antibodies, western blot samples were transferred onto PVDF membrane (Bio-Rad) in 20 mM Tris-Base, 200 mM glycine. Alexa fluorophore conjugated anti-rabbit or anti-mouse secondary antibodies (Molecular Probes) were used according to manufacturer's specification. Fluorescence was detected on a Typhoon 9400 Variable Mode Imager (Amersham Biosciences) and quantified using ImageJ (<http://rsb.info.nih.gov/ij/>). Actin antibodies were provided by J. Theriot (Stanford University) and anti-cyclin B was purchased from Santa Cruz Biotechnology.

In vitro binding of centromere proteins to chromatin arrays. Human and *Xenopus* CENP-C were *in vitro* translated (IVT) in rabbit reticulocyte extracts in the presence of 10 mCi ml⁻¹ [³⁵S]methionine (Perkin Elmer) using the TnT Quick-Coupled Transcription/Translation system (Promega) according to the manufacturer's instructions. For a binding reaction (60 µl total volume), 5 µl of each IVT protein were mixed with chromatin arrays in bead buffer (75 mM Tris-HCl pH 7.5, 50 mM NaCl, 0.25 mM EDTA, 0.05% Triton-X-100). The final nucleosome concentration per reaction was 60 nM. Reactions were incubated at

4 °C for 1 h. The beads were washed three times with bead buffer and resuspended in 4× SDS loading buffer. Samples were separated on a SDS–PAGE, Coomassie stained and after drying scanned using a phosphorimager (Typhoon 4200, Amersham Biosciences) and quantified using ImageJ (<http://rsb.info.nih.gov/ij/>). **Statistical analysis.** In each experiment, the relative levels of proteins associated with the chromatin arrays were normalized to values for wild type CENP-A arrays set to 1. For calculation of *P* values each data set was anchored at 1 and then log transformed followed by calculation of *P* values using a Student's *t*-test³⁷.

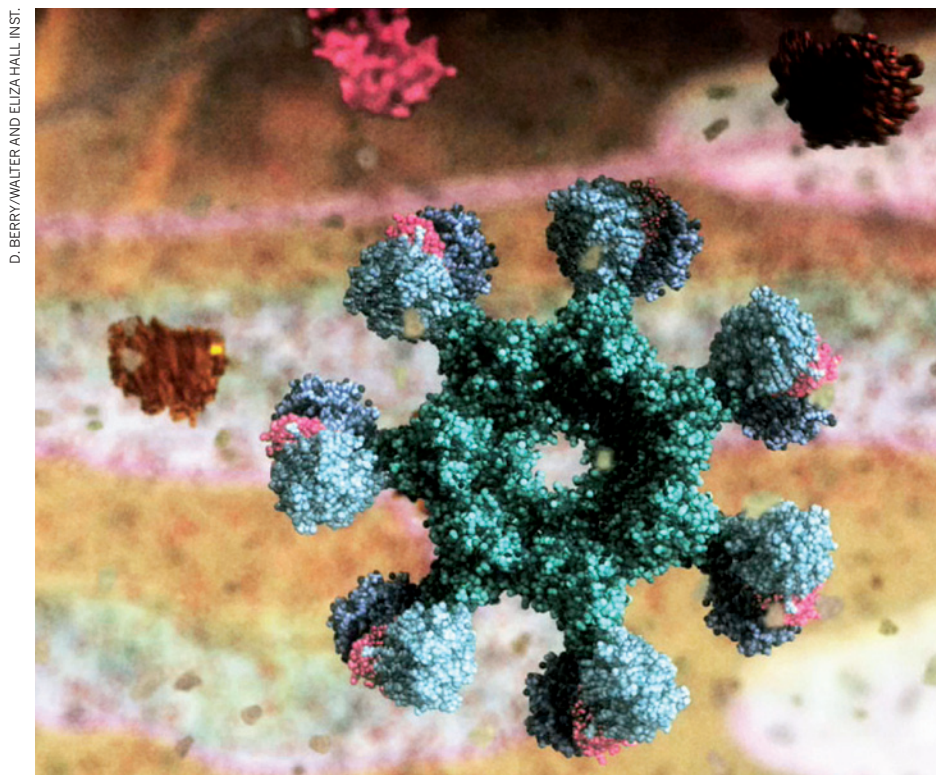
31. Luger, K., Rechsteiner, T. & Richmond, T. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304**, 3–19 (1999).
32. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).
33. Murray, A. W. Cell cycle extracts. *Methods Cell Biol.* **36**, 581–605 (1991).
34. Kim, S., Song, E., Lee, K. & Ferrell, J. Jr. Multisite M-phase phosphorylation of *Xenopus* Wee1A. *Mol. Cell. Biol.* **25**, 10580–10590 (2005).
35. Hannak, E. & Heald, R. Investigating mitotic spindle assembly and function *in vitro* using *Xenopus laevis* egg extracts. *Nature Protocols* **1**, 2305–2314 (2006).
36. Field, C. M., Oegema, K., Zheng, Y., Mitchison, T. J. & Walczak, C. E. Purification of cytoskeletal proteins using peptide antibodies. *Methods Enzymol.* **298**, 525–541 (1998).
37. Osborne, J. W. *Best Practices in Quantitative Methods* (Sage Publications, 2008).

CAREERS

TURNING POINT Postdoc's efforts bring explorer award and policy post **p.361**

ASK THE EXPERT Get incisive answers to career concerns go.nature.com/uuiugh

NATUREJOBS For the latest career listings and advice www.naturejobs.com



The protein structure known as the apoptosome, from the animation *Apoptosis* by Drew Berry.

BIOMEDICAL ILLUSTRATION

From monsters to molecules

Scientific animators are borrowing tools from Hollywood to breathe life into cells and molecules on screen.

BY CORIE LOK

Janet Iwasa listened as Samara Reck-Peterson struggled with a presentation. Reck-Peterson, a cell biologist, was trying to describe how dynein — a protein complex that acts as a molecular motor to transport cargo along the cell's cytoskeleton — moves. The complex is at the heart of Reck-Peterson's research, but she had only a styrofoam ball and pipe cleaners to demonstrate its complex actions at the faculty meeting. Iwasa, a

scientific animator and a lecturer in the same department as Reck-Peterson at Harvard Medical School in Boston, Massachusetts, recalls thinking: "We can do better than that."

So began a collaboration between the biologist and the animator. Over the past two years, the partnership has resulted in eye-catching images and animations of dynein that grace Reck-Peterson's presentation slides and her lab's website. Iwasa is now working on dynein animations that the researchers themselves can tinker with by manipulating the

motor's 'joints'. Reck-Peterson hopes that the animations will help her lab to design its next experiments, providing insight into exactly how this motor works. "The animations have made it easier to talk concretely about our ideas, both within the lab and with others in the field," says Reck-Peterson.

Iwasa is in the vanguard of scientific animators working in academia. Harvard's cell-biology department hired her three years ago to facilitate communication among faculty members and other scientists. Since then, she has worked with about a dozen researchers to visualize the molecules, pathways and cellular processes that they study, such as cell death. She also has a growing freelance business, 'onem micron illustration', creating animations, illustrations and websites for researchers at Harvard and other institutions.

Biomedical animators in the United States, Canada and elsewhere are seeing rising demand for their work from sectors including academic research, publishing, biotechnology and the drug industry. Animation studios have proliferated in the past five years, and medical-illustration master's-degree programmes have expanded their class sizes, with graduates generally able to find jobs with animation firms and research institutions. More and more scientists are seeking out animators, and a few, hoping to tinker with animation to aid their research rather than build a fully fledged career in it, are learning to use the tools themselves.

Driving this interest is an expansion in digital media connected with devices such as the iPad, and a burgeoning appreciation from publishers, scientists, educators, museum staff and others of the power of three-dimensional (3D) visualizations to communicate complex concepts. "The job is now getting the recognition and respect it deserves," says Drew Berry, a biomedical animator at the Walter and Eliza Hall Institute of Medical Research in Melbourne, Australia, who won a 'genius' grant from the MacArthur Foundation in Chicago, Illinois, last year.

TO ACADEMIA AND BEYOND

Scientific animators work with software similar to that used to create special effects and animated films in Hollywood, including a program called Autodesk Maya. But instead of creating monsters and explosions, they pull in data from a variety of sources, including review and research papers, to bring molecules and cells to life on screen. Dozens of papers can be necessary to inform a single animation. Animators also tap into scientific databases, ►

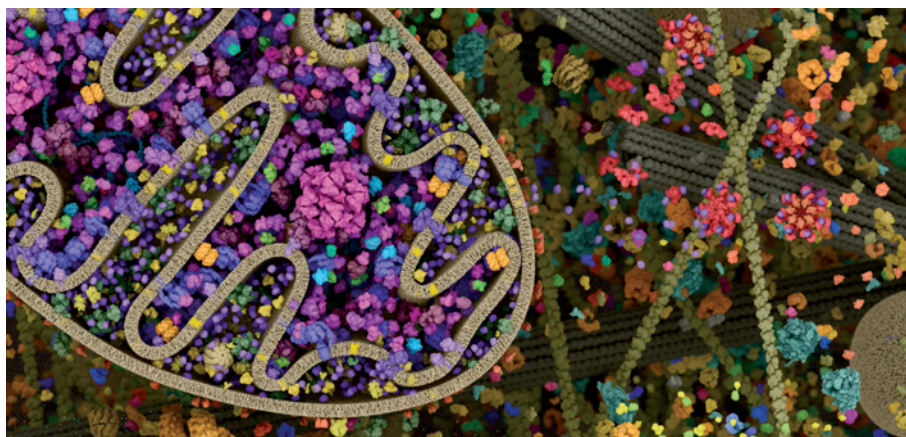
► such as the Protein Data Bank (www.pdb.org), to extract molecular structures, images, microscopy data and other key information. Some animators also spend a lot of time in close discussions with scientists to nail down details of how a molecule or cell moves or interacts with others. It is an iterative process, and a project can take months. “Diving into the latest data and making sense of it is challenging, but rewarding,” says Berry.

In academia, animators also develop courses and adapt animation tools for use in science. Eventually, they hope, visualizations could become a research tool used to develop, test and refine biomedical hypotheses, not just a method of communication. “As we get more data, we need better ways to synthesize the data into models,” says Iwasa.

Academia is only a small part of the market for scientific visualizations. Scientific animators can also be found at a growing number of commercial animation and design studios that specialize in biomedical work. These firms have been hiring increasing numbers of people over the past five to ten years, to serve an expanding client base. InViVo Communications in Toronto, Canada, started with three employees in 1998 and has since expanded to 46, including animators, programmers and sales people. XVIVO, a studio in Rocky Hill, Connecticut, has almost doubled its full-time staff to 15 employees over the past three years.

Such studios work mainly for the drug industry, but also for publishers, medical schools and teaching hospitals, and even for lawyers involved in malpractice lawsuits that require visuals as legal evidence. Medical-device, biotech and pharmaceutical companies use animations about their latest products in sales, marketing and educational materials. Visualizations also end up in museum exhibitions, classroom teaching tools, digital textbooks and documentaries, and on journal covers and websites.

According to 2009 data from a survey by the Association of Medical Illustrators in Lexington, Kentucky, illustrators and animators employed full-time earn a median salary of US\$52,000 at the start of their careers, \$65,000 in mid-career and up to \$150,000 as seasoned veterans. Many animators also work on a



Animations such as these images of mitochondria and proteins help researchers model cellular processes.

freelance basis, in which case their incomes can vary greatly; the median is \$79,000 a year, but incomes can reach up to \$250,000.

ARTFUL SCIENCE

Many scientific animators enjoy combining their passions for science, art and computers. Gaël McGill, director of molecular visualization at Harvard Medical School and president and chief executive of the studio Digizyme in Brookline, Massachusetts, spent his summers as a teenager with an aunt who was an art teacher, but he also loved science. He studied biology, art history and music as an undergraduate, then earned a PhD in cancer biology and completed a postdoc. Along the way, he discovered an interest in communicating science and started teaching himself how to use Maya at night, as a way to put his artistic skills to work.

Having a background in art or an eye for design and visual storytelling is crucial for scientific animators. An innate sense of aesthetics or some basic training in lighting, colour and composition to enable visual expression through drawing or other media is key to success, says Graham Johnson, an animator who will soon be starting a position at the University of California, San Francisco. He will be continuing his work developing software that, for example, integrates molecular modelling tools with animation programs to better connect raw scientific data with animation capabilities. “People with just a science bachelor’s and no formal art or illustration background will probably struggle,” he says.

Employers can quickly tell whether a budding animator has artistic talent by looking at their portfolio, or ‘demo reel’, which should showcase about a minute’s worth of animations, says Andrea Bielecki, president of InViVo. An animator’s personal website is also very telling — it should be easy to navigate, slick and quick to load demos.

A love of software and tinkering at the computer is essential. Computer-programming skills aren’t required, but they are in hot demand: interactive biomedical applications

for the web, or for the iPad and other devices, are the fastest-growing part of InViVo’s business, and skilled people are needed to write them, says Bielecki.

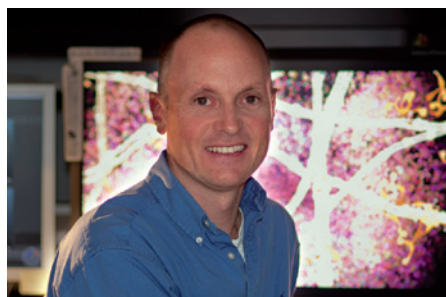
Although an artistic and production track record is paramount, having a scientific background can help with career advancement. Some design studios don’t require it, but McGill hires only people with scientific master’s degrees, PhDs or MDs to work at Digizyme. At the very least, those interested in scientific animation should have the ability to read and understand the relevant literature and to talk to scientists about their work.

A DEGREE OF ILLUSTRATION

Some in the field, such as McGill, are self-taught. Others learnt their skills through informal or part-time courses. Iwasa, for example, started her career as a graduate student in cell biology at the University of California, San Francisco. Already skilled in using programs such as Adobe Photoshop, Illustrator and Flash, she found that she enjoyed creating the figures that accompany manuscripts. With the support of her mentor, she took a basic animation course every Friday afternoon at another local university, and began making animations with her lab mates. After earning her PhD, she enrolled in a 12-week intensive course on how to use Maya.

Others attend accredited master’s programmes in medical illustration. There are only a handful of such courses in North America; one of these is the University of Toronto’s MSc in Biomedical Communications. The programme has a strong animation component and takes in 16 students each year, up from 8 in 2004 — a response to a rising number of applications and inquiries, and a growing job market. During the two-year professional programme, students take classes in biology and medicine, and learn about design concepts, software and the animation process. They also work closely with scientists on visualization projects. Many land jobs at animation studios, research institutions and elsewhere within a few months of graduation.

However, some graduates from the master’s programmes, especially from courses that



Medical animator Drew Berry has won a MacArthur grant, a sign of the field’s increasing recognition.

don't emphasize animation as much as that at Toronto, are not quite ready to work as production-level animators in a studio, says David Bolinsky, medical director of XVIVO. He recommends taking another year or so to attend a dedicated animation programme.

Many students entering the Toronto programme have bachelor of science degrees; an increasing number have advanced degrees, including PhDs, says Nicholas Woolridge, director of the course. Most are "passionate amateurs" in art, he adds. Applicants must submit a portfolio of work. "It doesn't have to be polished, but it needs to show that they can think visually and are visual problem-solvers," says Woolridge.

McGill, who teaches a one-year animation course for researchers at Harvard, is developing a new graduate programme, focusing on computer 'biovisualization'. Applicants would need to meet the same entrance requirements as for Harvard's other biology graduate courses, take the same first-year courses in cell and molecular biology and possibly even do lab rotations. McGill foresees his first set of students being existing PhD candidates who stay on for an extra year to earn an additional master's degree in visualization. McGill hopes to launch the master's programme in 2012, and eventually create a PhD programme. He sees his field shifting towards more research and software development, rather than just making animations.

Those interested in animation as a side activity or add-on skill to enhance their research can dabble at a basic level on their own. Many software packages, including Maya, can be downloaded at low cost, or even for free, as educational versions. "Just start playing and have fun," says Berry. Only then will people see whether they really enjoy the arduous and sometimes frustrating process of animation. Online tutorials, workshops and user forums can help amateur animators to learn on their own; see, for example, listings on www.molecularmovies.com. Most sites are geared towards film animation, but the concepts are the same.

McGill says that early-career scientists who master some skills in 3D animation can advance their careers by giving better seminars, using visual models of data to garner and inspire ideas and insights, and developing new research tools. In one case, a student of McGill's used his visualization skills to devise new DNA folding software that allows researchers to design their own molecules in three dimensions. "When you're going through the process of making a visualization," says McGill, "you come up with new questions and open up new ways of thinking." ■

Corie Lok is the editor of *Nature's Research Highlights*.

TURNING POINT

Jennifer Burney

Jennifer Burney, a physicist-turned-environmental-scientist at Scripps Institution of Oceanography at the University of California, San Diego, tells Nature about her upcoming tenure-track position in public policy and the unexpected honour of being named an Emerging Explorer by the National Geographic Society in Washington DC.

How did you end up pursuing both physics and international development?

I graduated with a bachelor's degree in history and science; I have always wanted to discover how science happens in a social context. But I enjoyed scientific research, which prompted me to pursue a physics PhD at Stanford University in California. I deferred graduate school for a year to volunteer with rebuilding efforts in Nicaragua following 1998's Hurricane Mitch. It was exciting to be in the field devising creative solutions. I eventually returned for my PhD, working to develop a superconducting camera that will help to capture images of cosmic bodies such as pulsars or exoplanets. But I continued to work for a non-profit group in Merced, California, called Engineers for a Sustainable World, which works with communities in the developing world.

You worked in the non-profit sector for a time, instead of going straight to a postdoc. Why?

As my PhD ended, I chose to try a non-academic route. My adviser said he would support me in whatever I decided. I knew that I wanted to investigate energy and climate issues in the developing world. So I cold-called a non-governmental organization (NGO), the Solar Electric Light Fund in Washington DC, which is involved in rural electrification around the world. One project was solar-powered drip irrigation in West Africa. They needed somebody to figure out how to evaluate the technology. That required assessing the design and how to make it cost-effective and sustainable.

How did this work influence your postdoc?

I continued working with the fund, and got interested in how energy and climate affect food security, water availability and agriculture. In 2008, I took a postdoc at Stanford's Program on Food Security and the Environment. Last October, I started a second postdoc at Scripps, where I began working on mitigating the climate impact of burning biomass for cooking and space heating. Now I'm involved in a project to replace cooking stoves with cleaner technologies over 100 square kilometres in northern India — and then



SCRIPPS INSTITUTION OF OCEANOGRAPHY

measuring the climate, health, hydrological and agricultural impacts over space and time.

Were you surprised to get a policy-based tenure-track position?

Yes. An advertisement for someone interested in science, technology, engineering and policy came up at the University of California, San Diego, and I thought, 'why not?' It was an exciting opportunity, not necessarily to straddle the worlds of NGOs and academia, but to have a job, starting next year, in which I would be teaching policy to scientists and science to policy-makers, while continuing my research.

How will the National Geographic Emerging Explorer distinction affect your career?

I'm still figuring it out. For the next year, *National Geographic* will track my scientific endeavours online. I just returned from the orientation meeting, where I met this year's class and previous explorers — and I have found a lot of common ground for collaboration. For example, one fellow works on ecological sanitation, and started a network of composting latrines in Haiti. We are planning some joint projects in West Africa, a region that needs new ways to generate fertilizer.

How have you benefited from stepping outside academia?

Leaving academia can invigorate your science. I'd encourage scientists to explore non-academic interactions — from giving public lectures to collaborating with NGOs. Being around non-scientists who channel their passion and understanding of science into real-life projects can shed light on how to make the most of your own expertise. ■

INTERVIEW BY VIRGINIA GEWIN

don't emphasize animation as much as that at Toronto, are not quite ready to work as production-level animators in a studio, says David Bolinsky, medical director of XVIVO. He recommends taking another year or so to attend a dedicated animation programme.

Many students entering the Toronto programme have bachelor of science degrees; an increasing number have advanced degrees, including PhDs, says Nicholas Woolridge, director of the course. Most are "passionate amateurs" in art, he adds. Applicants must submit a portfolio of work. "It doesn't have to be polished, but it needs to show that they can think visually and are visual problem-solvers," says Woolridge.

McGill, who teaches a one-year animation course for researchers at Harvard, is developing a new graduate programme, focusing on computer 'biovisualization'. Applicants would need to meet the same entrance requirements as for Harvard's other biology graduate courses, take the same first-year courses in cell and molecular biology and possibly even do lab rotations. McGill foresees his first set of students being existing PhD candidates who stay on for an extra year to earn an additional master's degree in visualization. McGill hopes to launch the master's programme in 2012, and eventually create a PhD programme. He sees his field shifting towards more research and software development, rather than just making animations.

Those interested in animation as a side activity or add-on skill to enhance their research can dabble at a basic level on their own. Many software packages, including Maya, can be downloaded at low cost, or even for free, as educational versions. "Just start playing and have fun," says Berry. Only then will people see whether they really enjoy the arduous and sometimes frustrating process of animation. Online tutorials, workshops and user forums can help amateur animators to learn on their own; see, for example, listings on www.molecularmovies.com. Most sites are geared towards film animation, but the concepts are the same.

McGill says that early-career scientists who master some skills in 3D animation can advance their careers by giving better seminars, using visual models of data to garner and inspire ideas and insights, and developing new research tools. In one case, a student of McGill's used his visualization skills to devise new DNA folding software that allows researchers to design their own molecules in three dimensions. "When you're going through the process of making a visualization," says McGill, "you come up with new questions and open up new ways of thinking." ■

Corie Lok is the editor of *Nature's Research Highlights*.

TURNING POINT

Jennifer Burney

Jennifer Burney, a physicist-turned-environmental-scientist at Scripps Institution of Oceanography at the University of California, San Diego, tells Nature about her upcoming tenure-track position in public policy and the unexpected honour of being named an Emerging Explorer by the National Geographic Society in Washington DC.

How did you end up pursuing both physics and international development?

I graduated with a bachelor's degree in history and science; I have always wanted to discover how science happens in a social context. But I enjoyed scientific research, which prompted me to pursue a physics PhD at Stanford University in California. I deferred graduate school for a year to volunteer with rebuilding efforts in Nicaragua following 1998's Hurricane Mitch. It was exciting to be in the field devising creative solutions. I eventually returned for my PhD, working to develop a superconducting camera that will help to capture images of cosmic bodies such as pulsars or exoplanets. But I continued to work for a non-profit group in Merced, California, called Engineers for a Sustainable World, which works with communities in the developing world.

You worked in the non-profit sector for a time, instead of going straight to a postdoc. Why?

As my PhD ended, I chose to try a non-academic route. My adviser said he would support me in whatever I decided. I knew that I wanted to investigate energy and climate issues in the developing world. So I cold-called a non-governmental organization (NGO), the Solar Electric Light Fund in Washington DC, which is involved in rural electrification around the world. One project was solar-powered drip irrigation in West Africa. They needed somebody to figure out how to evaluate the technology. That required assessing the design and how to make it cost-effective and sustainable.

How did this work influence your postdoc?

I continued working with the fund, and got interested in how energy and climate affect food security, water availability and agriculture. In 2008, I took a postdoc at Stanford's Program on Food Security and the Environment. Last October, I started a second postdoc at Scripps, where I began working on mitigating the climate impact of burning biomass for cooking and space heating. Now I'm involved in a project to replace cooking stoves with cleaner technologies over 100 square kilometres in northern India — and then



SCRIPPS INSTITUTION OF OCEANOGRAPHY

measuring the climate, health, hydrological and agricultural impacts over space and time.

Were you surprised to get a policy-based tenure-track position?

Yes. An advertisement for someone interested in science, technology, engineering and policy came up at the University of California, San Diego, and I thought, 'why not?' It was an exciting opportunity, not necessarily to straddle the worlds of NGOs and academia, but to have a job, starting next year, in which I would be teaching policy to scientists and science to policy-makers, while continuing my research.

How will the National Geographic Emerging Explorer distinction affect your career?

I'm still figuring it out. For the next year, *National Geographic* will track my scientific endeavours online. I just returned from the orientation meeting, where I met this year's class and previous explorers — and I have found a lot of common ground for collaboration. For example, one fellow works on ecological sanitation, and started a network of composting latrines in Haiti. We are planning some joint projects in West Africa, a region that needs new ways to generate fertilizer.

How have you benefited from stepping outside academia?

Leaving academia can invigorate your science. I'd encourage scientists to explore non-academic interactions — from giving public lectures to collaborating with NGOs. Being around non-scientists who channel their passion and understanding of science into real-life projects can shed light on how to make the most of your own expertise. ■

INTERVIEW BY VIRGINIA GEWIN

TWITTERSPACE

Careless talk.

BY WILLIAM MEIKLE

@Voyager2: I am currently 13 hrs 11 mins 26 secs of light-travel time from Earth

Dave was excited to find that he could follow the Voyager spacecraft on Twitter. He'd been obsessed with space, aliens and UFOs for as long as he could remember. He wanted to believe so bad, and being in touch with Voyager made him feel like he was reaching out into the vastness. In a small way, it felt like he was attempting *first contact*. His excitement soon turned to disappointment: the messages weren't coming from the craft but were being typed in by a nerd at NASA. It did however set him to thinking.

What if they're already here? What if they're watching us?

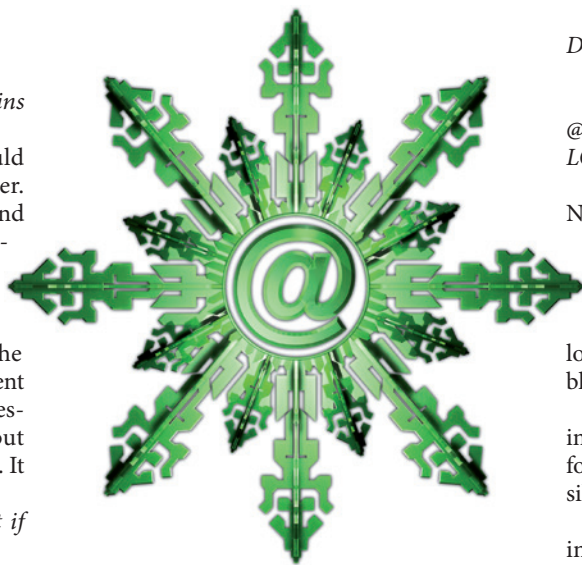
He did a search on Twitter — #aliens, #ufo and #invasion. The results were illuminating. @weegreenmen and @saucerzrus in particular shared many links, and many of them had nothing to do with aliens. What they *did* have a lot to do with was military infrastructure and economics for all the major powers on the planet. That was enough to make Dave think some more.

@weegreenmen: Check out Reuters. Big fluctuations in sterling today #invasion

He followed their tweets for several weeks. During that time he found out more than he needed to know about troop movements in Afghanistan, the North Korean nuclear programme, the perilous state of the Eurozone economies and, strangely, long-range weather forecasts for the Northern Hemisphere.

@saucerzrus: #ufo #aliens Major weather bomb in the Maritimes. Whoo-Hoo!

By now Dave was convinced he was on to something. The only way he would be able to find out what, was to join in on the conversation. He created a user on Twitter for the purpose. He spent a while looking for just the right name, and finally went with @littlegreybuddies. Then he needed a hook, to get their attention. The thing they were currently most interested in was weather patterns, so he started with that. He began by posting links to the North Atlantic storm watch sites, and actually found himself getting interested in the real-time tracking systems he found. That led him into ever more esoteric areas of research involving analyses of the movements of the jet stream, and apocalyptic warnings of serious trouble ahead for the world's climate.



@littlegreybuddies: Looks like the UK is in for a severe chill. So much for Global Warming #jetstream

That got their attention. He started to see his messages retweeted to the #aliens, #ufo and #invasion hash-tags. Slowly at first, he began retweeting messages posted by @weegreenmen and @saucerzrus, then started replying to their messages. In turn they started including him in their conversations, and seemed especially interested in his ongoing weather research.

@saucerzrus: #ufo #aliens Not long now till LUTZ #countdown

Buoyed by his acceptance Dave now felt that he had to do something to make sure he stayed there long enough to find out what was going on. He delved into university server systems and crept as close as he could to worldwide military information. From that he cobbled together a model of the coming month of where troops would be gathered, what the weather would be, and forecasted three weeks to come. He uploaded it all to a local ftp server and posted the link. Then he sat back and waited.

@saucerzrus: #ufo #aliens Hey @littlegreybuddies, THX man. #countdown brought forward

Dave was ecstatic. He'd made contact. He was now sure to the point of bursting with excitement that he was talking to actual alien entities here on Earth. He was on the verge of finding out their plan. And it never hurts to ask.

@littlegreybuddies: #countdown So when's

D-Day? More to the point ... what's D-Day?

The reply was almost immediate.

@saucerzrus: #countdown Watch the skies @littlegreybuddies Keep watching the skies LOL

Much to Dave's dismay things went quiet. Nobody posted to the hash-tags for a week, and @saucerzrus and @weegreenmen came back as discontinued when sent direct messages. He tried to force the issue by posting to the hash-tags, but he was a lone voice in the wilderness, tumbleweed blowing through his posts. No one replied.

Dave got desperate. He hacked his way into secure military installations, searching for a secret that would unlock his contacts' silence. He didn't find it.

What he did find was a growing disquiet in the military with the state of the upper atmosphere. Something was going on up there that had the top brass very worried. Dave was about to send a general tweet to see if anything was trending when he got a personal message. There was no sender identified, but he guessed who had sent it.

Time's up. Switch on the news.

He did as he was told.

"An unusual phenomenon is being reported all along the East Coast tonight. It is snowing in a zone stretching from New England to Labrador. Nothing unusual for this time of the year, except for the colour — the snow is green.

"Reports are also coming in that this snowfall is having strange effects on plant life in some areas. Scientists have taken samples of the substance for analysis, but as yet there is no official confirmation as to the cause of these events. All we can say for certain is that this is a deadly attack, from a source as yet unknown. FEMA has issued a preliminary statement asking people to remain indoors with doors and windows locked until the storm has passed, and we can only reiterate the importance of that advice. From what we have seen here, this country may never be the same again."

His laptop beeped.

A new tweet had just been posted to the hash-tags.

@saucerzrus: ROFLMAO Take us to your leader! PLS RT. #ufo #aliens #invasion #countdown = 0 ■

William Meikle is a Scottish writer resident in Canada with 10 novels published in the genre press and more than 200 short-story credits in 13 countries.

JACEY

The excitation of solar-like oscillations in a δ Sct star by efficient envelope convection

V. Antoci¹, G. Handler^{1,2}, T. L. Campante^{3,4}, A. O. Thygesen^{4,5}, A. Moya⁶, T. Kallinger^{1,7,8}, D. Stello⁹, A. Grigahcène³, H. Kjeldsen⁴, T. R. Bedding⁹, T. Lüftinger¹, J. Christensen-Dalsgaard⁴, G. Catanzaro¹⁰, A. Frasca¹⁰, P. De Cat¹¹, K. Uytterhoeven^{12,13,14,15}, H. Bruntt⁴, G. Houdek¹, D. W. Kurtz¹⁶, P. Lenz², A. Kaiser¹, J. Van Cleve¹⁷, C. Allen¹⁸ & B. D. Clarke¹⁷

Delta Scuti (δ Sct) stars are opacity-driven pulsators with masses of $1.5\text{--}2.5M_{\odot}$, their pulsations resulting from the varying ionization of helium. In less massive stars² such as the Sun, convection transports mass and energy through the outer 30 per cent of the star and excites a rich spectrum of resonant acoustic modes. Based on the solar example, with no firm theoretical basis, models predict that the convective envelope in δ Sct stars extends only about 1 per cent of the radius³, but with sufficient energy to excite solar-like oscillations^{4,5}. This was not observed before the Kepler mission⁶, so the presence of a convective envelope in the models has been questioned. Here we report the detection of solar-like oscillations in the δ Sct star HD 187547, implying that surface convection operates efficiently in stars about twice as massive as the Sun, as the *ad hoc* models predicted.

Thirty days of continuous observations of HD 187547 (KIC 7548479) by the Kepler mission with a cadence of 1 min led to its identification as a δ Sct pulsator (Fig. 1a, b). In contrast to the non-uniformly distributed signals at low frequencies, the observed regularly spaced peaks at high frequencies (Fig. 1c) suggest that we also observe high-radial-order overtones as expected for stochastically excited solar-like oscillations. For such oscillations the observed comb-like frequency structure (with the large frequency separation $\Delta\nu$ indicating the frequency separation between consecutive radial overtones of like degree) is the result of mainly radial and dipolar pulsation modes, whereas for δ Sct stars it is not clear which modes are excited to observable amplitudes. The strikingly broadened structures observed only at high frequencies (Figs 1f and 2b, c) suggest that each is due either to single damped and stochastically re-excited oscillations or to very close unresolved frequencies of coherent oscillations.

Here we use spectroscopic observations to derive an effective temperature $T_{\text{eff}} = 7,500 \pm 250$ K, a surface gravity of $\log g = 3.90 \pm 0.25$ dex (c.g.s.) and a projected rotational velocity of $v \sin i = 10.3 \pm 2.3$ km s⁻¹ (see Supplementary Information for details). We identify HD 187547 as an Am star from chemical element abundance analysis, which is consistent with the observed low $v \sin i$ typical for these stars. Am stars are stars of spectral type A showing atmospheric underabundance when compared with the Sun in the chemical elements Sc and Ca, and an overabundance of Ba, Sr and Y (ref. 7). We compute a photospheric metallicity (all elements except H and He) of $Z = 0.017$, which is larger than the solar value of $Z = 0.0134$ (ref. 8).

About two-thirds of Am stars are primary components of spectroscopic binary systems⁹. The Am phenomenon is connected to slow

rotation, which is not common in A type stars. Binarity is believed to act as a braking mechanism slowing down the rotation and allowing spectral peculiarities to occur as a result of element diffusion¹⁰. Pulsating Am stars still represent a challenge to theory, because He is expected to settle gravitationally and should only partly be present in the He II ionization zone where the δ Sct pulsations are excited. In other words, theoretical models predict that the hottest and youngest A-type stars should not pulsate¹⁰, which is in contradiction with recent observations¹¹. As the stars evolve, their convective envelopes deepen and efficiently mix the stellar matter, erasing the observed chemical peculiarities in the atmospheres, allowing the opacity mechanism to drive pulsation in the He II ionization zone. Using the observed solar-like oscillations reported here, the depth of the convective envelope can be derived (hence the mixing length), probing the diffusion of He and heavy elements in this star. This will contribute significantly to revising the interaction between pulsation and diffusion in models of Am stars.

Seven radial velocity measurements of HD 187547, spread over 153 days, give no evidence for a short-period binary system. In addition, the absence of any detectable contribution by a potential close companion to the spectrum implies a considerably less luminous star of spectral type G or later. The expected amplitudes and frequency of maximum oscillation power for such a star are inconsistent with the observations, leading to the conclusion that the signal observed in Fig. 1c cannot originate from a companion. The observed amplitude spectrum of HD 187547 is not affected by a background star because the fraction of light in the aperture from neighbouring stars is only 1.5%. Other chemically peculiar pulsating stars situated, as the δ Sct stars, in the classical instability strip in the Hertzsprung–Russell diagram¹² are the rapidly oscillating Ap stars. Their high-radial-order pulsation modes are triggered by the opacity mechanism acting in the hydrogen ionization zone, often showing equidistant multiplets in the frequency spectrum as a result of the alignment of the pulsation axes with strong magnetic fields¹³. The strong magnetic fields as seen in rapidly oscillating Ap stars are, however, not observed in Am stars¹⁴. We therefore exclude the possibility that HD 187547 is a hybrid of a δ Sct and a rapidly oscillating Ap star.

In Fig. 3 we show an échelle diagram comparing the observed frequencies with a model of a star similar to HD 187547, demonstrating again the clear structures separated by $\Delta\nu$ at high frequencies and the non-structured distribution at lower frequencies. For the high-frequency modes we derive a mean large frequency separation $\Delta\nu$ of 40.5 ± 0.6 μ Hz. Using the empirical relation¹⁵

¹Institute of Astronomy, University of Vienna, Türkenschanzstraße 18, A-1180 Vienna, Austria. ²Copernicus Astronomical Center, Bartycycka 18, 00-716 Warsaw, Poland. ³Centro de Astrofísica, Departamento de Física e Astronomia-Faculdade de Ciências, Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal. ⁴Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark. ⁵Nordic Optical Telescope, Apartado 474, E-38700 Santa Cruz de La Palma, Santa Cruz de Tenerife, Spain. ⁶Departamento de Astrofísica, Centro de Astrobiología, Instituto Nacional de Técnica Aeroespacial – Consejo Superior de Investigaciones Científicas, PO Box 78, E-28691, Villanueva de la Cañada, Madrid, Spain. ⁷Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, British Columbia V6T 1Z1, Canada. ⁸Instituut voor Sterrenkunde, University of Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium. ⁹Sydney Institute for Astronomy (SIfA), School of Physics, University of Sydney, New South Wales 2006, Australia. ¹⁰Istituto Nazionale di Astrofisica – Osservatorio Astrofisico di Catania, via S. Sofia 78, I-95123 Catania, Italy. ¹¹Royal Observatory of Belgium, Ringlaan 3, B-1180 Brussels, Belgium. ¹²Laboratoire AIM, CEA/DSM-CNRS Université Paris Diderot, Commissariat à l'Energie Atomique et aux Energies Alternatives, Institut de Recherche sur Les Lois Fondamentales, Service d'Astrophysique, Centre de Saclay, F-91191, Gif-sur-Yvette, France. ¹³Kiepenheuer-Institut für Sonnenphysik, Schöneckstrasse 6, 79104 Freiburg, Germany. ¹⁴Instituto de Astrofísica de Canarias, 38200 La Laguna, Tenerife, Spain. ¹⁵Departamento de Astrofísica, Universidad de La Laguna, 38205 La Laguna, Tenerife, Spain. ¹⁶Jeremiah Horrocks Institute, University of Central Lancashire, Preston PR1 2HE, UK. ¹⁷SETI Institute/NASA Ames Research Center, Moffett Field, California 94035, USA. ¹⁸Orbital Sciences Corporation/NASA Ames Research Center, Moffett Field, California 94035, USA.

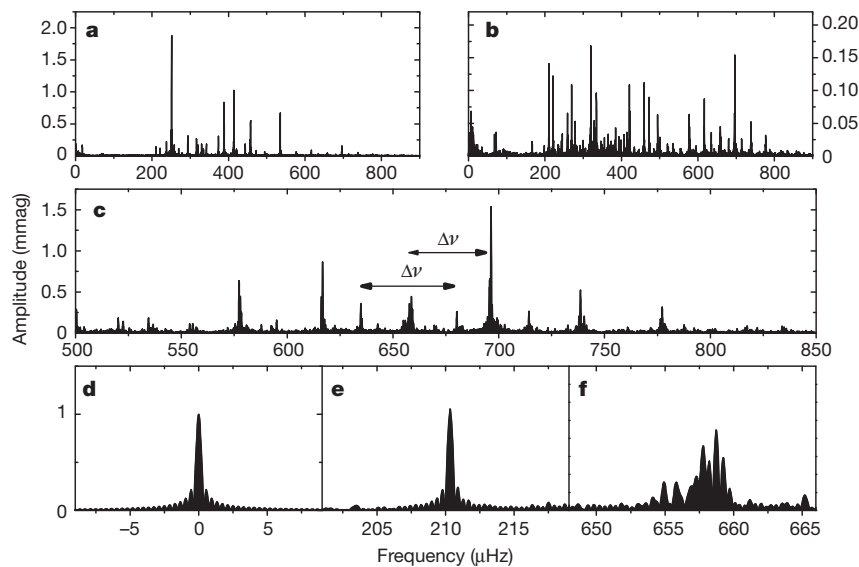


Figure 1 | Fourier amplitude spectra of the Kepler light curve of HD 187547. **a**, Fourier spectrum covering the entire frequency range in which significant signals were observed with a dominant frequency at 251 μHz and an amplitude of 2 mmag, typical for a δ Sct star. **b**, The multimode oscillations of HD 187547 are shown by subtracting 16 sinusoids corresponding to the most prominent oscillations, revealing a large number of additional significant frequencies. **c**, The region between 500 and 850 μHz shows a clear pattern of roughly equally spaced peaks, which we interpret as high-order consecutive radial overtones. The comb-like structure expected for high-order radial

overtones is clearly visible. The broadened peaks suggest damped/re-excited solar-like oscillations. The black arrows denoted $\Delta\nu$ indicate the large separation between consecutive radial and dipole modes. **d**, Spectral window. The shape of the window function is defined by the length and sampling of the data set. Any coherent signal will have the same profile. **e**, Example for one of the modes driven by the opacity mechanism in HD 187547. **f**, A supposed solar-like oscillation mode observed in HD 187547, displaying a broadened structure suggestive of a short mode lifetime.

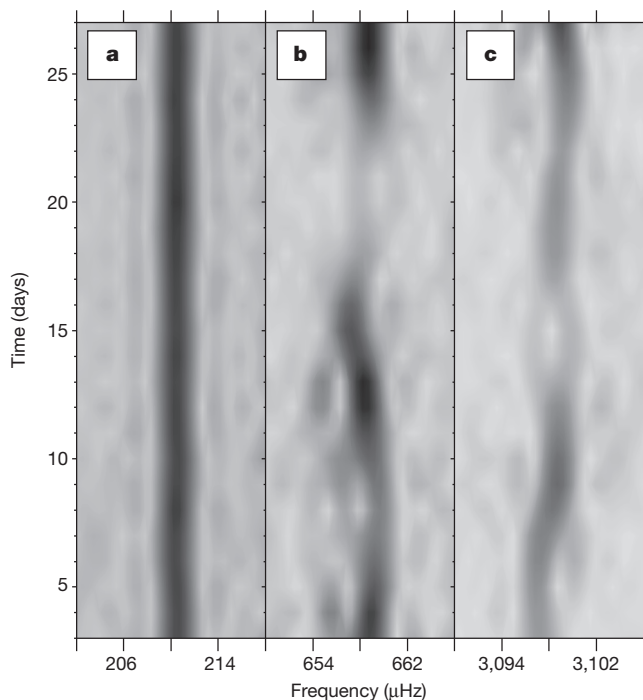


Figure 2 | Time-Fourier spectrum. Here we again highlight the difference in temporal variability between the modes interpreted as stochastic modes and the coherent, opacity-driven peaks at low frequencies. The time-fourier spectrum was computed with a running filter of full-width at half-maximum = 5 days, comparable to the mean mode lifetime. **a**, An opacity-driven mode (the same as in Fig. 1e) showing temporal stability in the δ Sct frequency region. **b**, Stochastic mode observed in HD 187547, showing an erratic behaviour as expected for solar-like oscillations (the same as in Fig. 1f). **c**, For comparison, a stochastic oscillation mode observed in the Sun. The solar data were obtained from the SOHO VIRGO instrument. The data set has the same length and sampling as for HD 187547; that is, 30 days and 1 min, respectively. Further details of frequency analyses and tests on artificial data sets (Supplementary Fig. 1) to verify our interpretation are in the Supplementary Information.

$\Delta\nu = (0.263 \pm 0.009) \mu\text{Hz} (\nu_{\text{max}} \mu\text{Hz}^{-1})^{0.772 \pm 0.005}$ we obtain a frequency of maximum power $\nu_{\text{max}} = 682_{-43}^{+41} \mu\text{Hz}$. This is in very good agreement with the highest-amplitude mode in the supposed stochastic frequency region at 696 μHz . The possibility that what we observe is $0.5\Delta\nu$ in the frequency spectrum is ruled out because this would require a ν_{max} at about 1,673 μHz , where no signal is observed. We can also exclude the observation of $2\Delta\nu$ because that would place ν_{max} at 277 μHz , close to the dominant opacity-driven mode at 251 μHz .

The amplitudes of solar-like oscillations are determined by the interaction between driving and damping defined by different physical processes², such as modulation of the turbulent momentum and heat fluxes by pulsation. The exact contribution to driving and damping by each of these processes is still not well understood, resulting in uncertainties in the predictions of the stochastically excited mode amplitudes¹⁶, particularly in hotter stars^{2,3} in which the convective envelopes are shallow. We expect the mixing length, the amplitudes and mode lifetimes to constrain the anisotropy of the convective velocity field, parameters that all semi-analytical convection models rely on¹⁷.

For HD 187547 we measure a peak-amplitude per radial mode¹⁸ for the assumed stochastic signal of 56 ± 2 p.p.m., which after bolometric correction¹⁹ results in 67 ± 3 p.p.m. (see Supplementary Information for details). From the empirical scaling relation²⁰ and using a bolometric solar peak-amplitude of 3.6 p.p.m. (ref. 21) we obtain a predicted peak amplitude of $A = 14 \pm 9$ p.p.m. The mean mode lifetimes are measured²² as 5.7 ± 0.8 days. Empirical relations predict a mode lifetime for a star with $T_{\text{eff}} = 7,500 \pm 250$ K of the order of one day²³ or shorter²⁴, which is not in agreement with what we measure for HD 187547. However, these scaling relations (for amplitude and mode lifetimes) are based on few observed stars, and none of them is calibrated in the temperature domain of our target, for which the physical conditions in the convection zone are expected to be very different. Furthermore, given that HD 187547 is metal overabundant in comparison with the Sun, the observed amplitude is expected to be higher^{3,25} than predicted from simple scaling, which is indeed the case. The power of a mode is directly proportional to the mode lifetime provided that the energy supply rate over the mode inertia is constant²⁶, which further supports

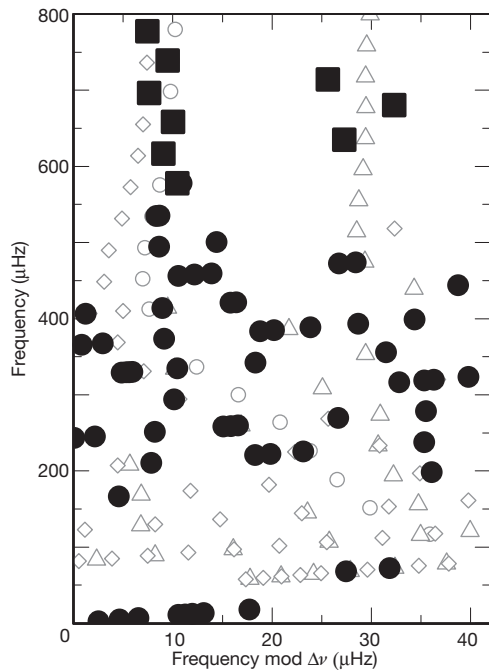


Figure 3 | Échelle diagram of HD 187547. Here we plot 69 extracted frequencies as a function of frequency modulo the large separation ($\Delta\nu = 40.5 \mu\text{Hz}$). Frequencies equally spaced by $\Delta\nu$ will form vertical ridges in the échelle diagram. To guide the eye, we show theoretically predicted frequencies of pulsation modes for a $1.85M_{\odot}$ stellar model. Ridges of $l = 0$ modes are represented by open circles, $l = 1$ by open triangles and $l = 2$ by open diamonds. Detailed modelling of the star is beyond the scope of this paper. The supposed solar-like modes (filled squares) between 500 and 870 μHz show clear ridges, as expected for high-order acoustic oscillations similar to what is observed in solar-like stars. The lower frequencies that we attribute to δ Sct pulsation (filled circles) excited by the opacity mechanism show no obvious regular patterns.

the higher amplitude because the observed mode lifetimes are also longer than expected. An additional factor that is not considered in any scaling relations is the chemical peculiarity of our target. In summary, these factors make HD 187547 an intriguing case for further theoretical analyses of stochastic oscillations and the potential interaction with the opacity mechanism in δ Sct stars.

The amplitude distribution for stochastic pulsation can be described as a Rayleigh distribution, provided that the examined time series are much shorter than the mode lifetimes. The relation between the mean amplitude $\langle A \rangle$ and its standard deviation $\sigma(A)$ can then be written as²⁷ $(4/\pi - 1)^{0.5} \langle A \rangle \approx 0.52 \langle A \rangle$. This is not valid for opacity-driven pulsation. For HD 187547 we therefore expect to obtain two different regimes of the ratio $\sigma(A)/\langle A \rangle$ for the two groups of oscillation modes (Supplementary Fig. 2). Indeed, we see that the δ Sct frequencies have a lower value of $\sigma(A)/\langle A \rangle$ than the supposed solar-like modes, giving further evidence for the stochastic nature of the latter (see Supplementary Information for details).

We cannot strictly exclude the possibility that the signals between 578 and 868 μHz are due to unresolved modes of pulsation excited by the opacity mechanism, because high-radial-order acoustic modes can also be observed in hot δ Sct stars. Nevertheless, as shown in Fig. 1 this would imply that δ Sct pulsation covers the region between 205 and 870 μHz continuously. According to current theory, the opacity mechanism acting in the He II ionization zone cannot excite modes spanning 16 radial orders for a star with parameters like those of HD 187547 (ref. 28). Further support for the discovery of solar-like oscillations comes from spectroscopic observations that also indicate the presence of convective motions in the atmospheres of A and Am stars²⁹. In addition, signatures of granulation noise in δ Sct stars have been

reported from photometric measurements³. Opacity-driven pulsations are also observed in more massive stars ($8\text{--}16M_{\odot}$), known as β Cephei stars (in this case the opacity mechanism acts in the ionization region of the iron-group elements). The unexpected detection of solar-like oscillations in such a star³⁰ (with a mass of $10M_{\odot}$) suggests that both types of pulsation, opacity-driven and stochastically excited, can coexist and can have overlapping frequency domains. The similar timescales of the different oscillation types imply a possible interaction between the two mechanisms.

Received 7 April; accepted 26 July 2011.

Published online 14 September 2011.

- Breger, M. in *Delta Scuti and Related Stars* (ASP Conf. Ser. Vol. 210) (eds Breger, M. & Montgomery, M.) 3–42 (Astronomical Society of the Pacific, 2000).
- Chaplin, W. J. et al. Ensemble asteroseismology of solar-type stars with the NASA Kepler mission. *Science* **332**, 213–216 (2011).
- Kallinger, T. & Matthews, J. M. Evidence for granulation in early A-type stars. *Astrophys. J.* **711**, L35–L39 (2010).
- Houdek, G., Balmforth, N. J., Christensen-Dalsgaard, J. & Gough, D. O. Amplitudes of stochastically excited oscillations in main-sequence stars. *Astron. Astrophys.* **351**, 582–596 (1999).
- Samadi, R., Goupil, M.-J. & Houdek, G. Solar-like oscillations in delta Scuti stars. *Astron. Astrophys.* **395**, 563–571 (2002).
- Koch, D. G. et al. Kepler mission design, realized photometric performance, and early science. *Astrophys. J.* **713**, L79–L86 (2010).
- Preston, G. W. The chemically peculiar stars of the upper main sequence. *Annu. Rev. Astron. Astrophys.* **12**, 257–277 (1974).
- Asplund, M., Grevesse, N., Sauval, A. J. & Scott, P. The chemical composition of the Sun. *Annu. Rev. Astron. Astrophys.* **47**, 481–522 (2009).
- Carquillat, J.-M. & Prieur, J.-L. Contribution to the search for binaries among Am stars. VIII. New spectroscopic orbits of eight systems and statistical study of a sample of 91 Am stars. *Mon. Not. R. Astron. Soc.* **380**, 1064–1078 (2007).
- Turcotte, S., Richer, J., Michaud, G. & Christensen-Dalsgaard, J. The effect of diffusion on pulsations of stars on the upper main sequence— δ Scuti and metallic A stars. *Astron. Astrophys.* **360**, 603–616 (2000).
- Balona, L. et al. Kepler observations of Am stars. *Mon. Not. R. Astron. Soc.* **414**, 792–800 (2011).
- Handler, G. Confirmation of simultaneous p and g mode excitation in HD 8801 and γ Peg from time-resolved multicolour photometry of six candidate ‘hybrid’ pulsators. *Mon. Not. R. Astron. Soc.* **398**, 1339–1351 (2009).
- Kurtz, D. W. Rapidly oscillating AP stars. *Mon. Not. R. Astron. Soc.* **200**, 807–859 (1982).
- Aurière, M. et al. No detection of large-scale magnetic fields at the surfaces of Am and HgMn stars. *Astron. Astrophys.* **523**, A40, doi:10.1051/0004-6361/201014848 (2010).
- Stello, D., Chaplin, W. J., Basu, S., Elsworth, Y. & Bedding, T. R. The relation between $\Delta\nu$ and ν_{max} for solar-like oscillations. *Mon. Not. R. Astron. Soc.* **400L**, L80–L84 (2009).
- Houdek, G. Solar-type variables. *AIP Conf. Proc.* **1170**, 519–530 (2009).
- Samadi, R., Belkacem, K., Goupil, M.-J., Kupka, F. & Dupret, M.-A. Solarlike oscillation amplitudes and line-widths as a probe for turbulent convection in stars. *Proc. Int. Astron. Un.* **239**, 349–357 (2007).
- Kjeldsen, H. et al. The amplitude of solar oscillations using stellar techniques. *Astrophys. J.* **682**, 1370–1375 (2008).
- Ballot, J., Barban, C. & van’t Veer-Menneret, C. Visibilities and bolometric corrections for stellar oscillation modes observed by Kepler. *Astron. Astrophys.* **531**, A124, doi:10.1051/0004-6361/201016230 (2011).
- Kjeldsen, H. & Bedding, T. R. Amplitudes of solar-like oscillations: a new scaling relation. *Astron. Astrophys.* **529**, L8, doi:10.1051/0004-6361/201116789 (2011).
- Michel, E. et al. Intrinsic photometric characterisation of stellar oscillations and granulation. *Astron. Astrophys.* **495**, 979–987 (2009).
- Gruberbauer, M., Kallinger, T., Weiss, W. W. & Guenther, D. B. On the detection of Lorentzian profiles in a power spectrum: a Bayesian approach using ignorance priors. *Astron. Astrophys.* **506**, 1043–1053 (2009).
- Chaplin, W. J., Houdek, G., Karoff, C., Elsworth, Y. & New, R. Mode lifetimes of stellar oscillations. Implications for asteroseismology. *Astron. Astrophys.* **500**, L21–L24 (2009).
- Baudin, F. et al. Amplitudes and lifetimes of solar-like oscillations observed by CoRoT. Red-giant versus main-sequence stars. *Astron. Astrophys.* **529**, A84, doi:10.1051/0004-6361/201014037 (2011).
- Samadi, R., Ludwig, H.-G., Belkacem, K., Goupil, M. J. & Dupret, M.-A. The CoRoT target HD 49933. I. Effect of the metal abundance on the mode excitation rates. *Astron. Astrophys.* **509**, A15, doi:10.1051/0004-6361/200911867 (2010).
- Chaplin, W. J. et al. On model predictions of the power spectral density of radial solar p modes. *Mon. Not. R. Astron. Soc.* **360**, 859–868 (2005).
- Chang, H.-Y. & Gough, D. O. On the power distribution of solar p modes. *Solar Phys.* **181**, 251–263 (1998).
- Pamyatnykh, A. A. in *Delta Scuti and Related Stars* (ASP Conf. Ser. Vol. 210) (eds Breger, M. & Montgomery, M.) 215–246 (Astronomical Society of the Pacific, 2000).

29. Landstreet, J. D. *et al.* Atmospheric velocity fields in tepid main sequence stars. *Astron. Astrophys.* **503**, 973–984 (2009).
30. Belkacem, K. *et al.* Solar-like oscillations in a massive star. *Science* **324**, 1540–1542 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the entire Kepler team, without whom these results would not be possible. V.A., G.Ha. and G.Ho. were supported by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung. V.A. also thanks L. Fossati for his help. A.M. acknowledges the funding of AstroMadrid, who was also supported by Spanish grants. T.R.B. and D.S. acknowledge support from the Australian Research Council. T.L. was supported by the Austrian Agency for International Cooperation in Education and Research. K.U. acknowledges financial support from the Deutsche Forschungsgemeinschaft. Funding for this Discovery mission is provided by NASA's Science Mission Directorate.

Author Contributions V.A. discovered the star among the Kepler targets, analysed it and found the solar-like oscillations (as a part of her PhD thesis), did spectroscopic analyses, frequency analyses, the test on the stochastic nature of the signal, and interpretations, and wrote the paper. G.Ha. had the idea for this project and supervised V.A., helped with analyses, interpretations and writing the paper. T.L.C. contributed to the analyses of the stochastic modes and also to the test on the stochastic nature of the signal. A.O.T. observed the target spectroscopically at Nordic Optical Telescope, identified the star as

an Am star and did spectroscopic analyses. A.M. contributed to the statistical test on the nature of the stochastic signal. T.K. helped interpretations, data analyses and writing the paper, and delivered the mode lifetimes. D.S. helped with data analyses, writing the paper and made Fig. 2. A.G. helped with theoretical interpretations and the time-Fourier analyses. T.R.B. helped with interpretations and writing the paper. H.K. contributed to the analyses, also by supervising V.A., and is a member of the Kepler Asteroseismic Investigation Steering Committee. J.C.-D. helped with the theoretical support, writing the paper and is a member of the Kepler Asteroseismic Investigation Steering Committee. T.L. confirmed the Am identification, excluding the Ap character of the star. G.C., A.F. and A.K. did spectroscopic analyses. P.D.C. was Principal Investigator and observer for the spectroscopic data from McDonald observatory. K.U. was Co-Investigator of the McDonald data and coordinated the ground-based observations. H.B. was Principal Investigator for the observations with the NARVAL spectrograph and did spectroscopic analyses. G.Ho. and P.L. helped with theoretical interpretations and writing the paper. D.W.K. helped with the Am classification and writing the paper and is leader of the delta Scuti working group of the Kepler Asteroseismic Science Consortium. J.V.C., C.A. and B.D.C. are part of the Kepler team and were involved in designing and operating the satellite. All co-authors contributed to discussions and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to V.A. (victoria.antoci@univie.ac.at).

Synthetic chromosome arms function in yeast and generate phenotypic diversity by design

Jessica S. Dymond^{1,2†}, Sarah M. Richardson^{1,3}, Candice E. Coombes^{1,2}, Timothy Babatz^{1,3}, H  lo  se Muller⁴, Narayana Annaluru⁴, William J. Blake^{5†}, Joy W. Schwerzmann^{4†}, Junbiao Dai^{1,2†}, Derek L. Lindstrom^{6†}, Annabel C. Boeke^{1†}, Daniel E. Gottschling⁶, Srinivasan Chandrasegaran⁴, Joel S. Bader^{1,7} & Jef D. Boeke^{1,2}

Recent advances in DNA synthesis technology have enabled the construction of novel genetic pathways and genomic elements, furthering our understanding of system-level phenomena^{1–7}. The ability to synthesize large segments of DNA allows the engineering of pathways and genomes according to arbitrary sets of design principles. Here we describe a synthetic yeast genome project, Sc2.0, and the first partially synthetic eukaryotic chromosomes, *Saccharomyces cerevisiae* chromosome synIXR, and semi-synVIL. We defined three design principles for a synthetic genome as follows: first, it should result in a (near) wild-type phenotype and fitness; second, it should lack destabilizing elements such as tRNA genes or transposons^{8,9}; and third, it should have genetic flexibility to facilitate future studies. The synthetic genome features several systemic modifications complying with the design principles, including an inducible evolution system, SCRaMbLE (synthetic chromosome rearrangement and modification by loxP-mediated evolution). We show the utility of SCRaMbLE as a novel method of combinatorial mutagenesis, capable of generating complex genotypes and a broad variety of phenotypes. When complete, the fully synthetic genome will allow massive restructuring of the yeast genome, and may open the door to a new type of combinatorial genetics based entirely on variations in gene content and copy number.

The first phase of any genome engineering project is design (Supplementary Text 1). We designed the right arm of chromosome IX (IXR) according to the three principles outlined above and in Box 1. IXR is the smallest chromosome arm in the genome and features several genomic elements of interest (Fig. 1a), making it suitable for a pilot study. The designed sequence, synIXR, is based on a native IXR sequence extending from open reading frame (ORF) *YIL002W* through the centromere and the remainder of chromosome IXR, an 89,299-base-pair (bp) sequence (native IXR position 350,585–438,993 (ref. 10)). In accordance with the second design principle, a transfer RNA gene, a Ty1 long terminal repeat (LTR), and telomeric sequences were removed. The final synIXR sequence, 91,010 bp, is slightly longer than the native sequence owing to the inclusion of 43 loxPsym sites, and it replaces 20.3% of the native chromosome. A 30-kilobase (kb) telomeric segment of the left arm of chromosome VI (semi-synVIL) was similarly designed (Fig. 1b and Supplementary Text 2), and replaced 15.7% of the native chromosome. Of the original sequence lengths, 17% was changed by base substitution, deleted, or inserted during design of the two synthetic segments (Supplementary Table 1). Sequences were submitted to GenBank (sequences synIXR:JN020955 and semi-synVIL:JN020956 are also available in Supplementary Information).

We systematically introduced two sets of changes *in silico* using the genome editing suite BioStudio (S.M.R., J.S.D., J.D.B. and J.S.B., unpublished data): TAG/TAA stop-codon swaps and PCRTags sequences (see Supplementary Text 1). In recognition of the third design principle, the elimination of the TAG stop codon by recoding to TAA frees a codon for future expansion of the genetic code (for example, by adding a twenty-first, unnatural amino acid^{11,12}), and could serve as a future mechanism of reproductive isolation and control. PCRTags are short pairs of recoded sequences, unique to either the wild-type or synthetic genome. They serve as convenient, low-cost, closely spaced genetic markers for verifying the introduction of synthetic sequence and the removal of native sequence by allowing the design of PCR primers for rapid evaluation of the presence of synthetic sequences and absence of native sequences. This is critical for evaluating the incorporation of synthetic DNA (see below and Supplementary Text 2). PCRTags, designed *in silico*, were tested in triplicate to verify specificity (Supplementary Fig. 1 and Supplementary Tables 2 and 3).

LoxPsym sequences are nondirectional loxP sites that are capable of recombining in either orientation¹³. Theoretically, they produce inversions or deletions with equal probability. Under the third design principle, these sites form the substrate for the inducible SCRaMbLE system and are intended to generate combinatorial diversity. We inserted loxPsym sites 3 bp after the stop codon of each nonessential gene and at major landmarks, such as sites of LTR and tRNA deletions, flanking the centromere *CEN9*, and adjacent to telomeres (Fig. 1 and Supplementary Text 1). LoxPsym sites inserted at equivalent positions genome-wide will allow the formation of many structurally distinct genomes.

After completion of chromosome design and construction, ‘arm-swap’ strains, wherein the wild-type sequence was replaced with synthetic sequence, were generated. The synIXR chromosome, cloned in a circular bacterial artificial chromosome (BAC) vector, includes all sequences needed for propagation in yeast and bacteria (Fig. 1a). We introduced synIXR into a diploid strain by transformation (Fig. 2a); typically, about 10–15% of the synIXR transformants obtained were positive for all PCRTags pairs tested (Fig. 2d). We chose one such transformant, strain A (Fig. 2a), and truncated one native IXR homologue (IX  R) by transforming with a suitably designed linear DNA fragment¹⁴, introducing a selectable marker (*URA3*) and a telomere seed sequence, generating strain C (Fig. 2b). Chromosome truncation was confirmed by pulsed-field gel electrophoresis analysis (Fig. 2c), and strain C was sporulated to generate haploids carrying synIXR and

¹High Throughput Biology Center, Johns Hopkins University School of Medicine, 733 North Broadway, Baltimore, Maryland 21205, USA. ²Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA. ³McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 733 North Broadway, Baltimore, Maryland 21205, USA. ⁴Bloomberg School of Public Health, Johns Hopkins University, 615 North Wolfe Street, Baltimore, Maryland 21205, USA. ⁵Codon Devices, 99 Erie Street, Cambridge, Massachusetts 02139, USA. ⁶Fred Hutchinson Cancer Research Center, Mailstop A3-025, PO Box 19024, Seattle, Washington 98109, USA. ⁷Department of Biomedical Engineering, Whiting School of Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA. †Present addresses: Animal Biosciences and Biotechnology Laboratory, Agricultural Research Service, USDA, 10300 Baltimore Avenue, Beltsville, Maryland 20705, USA (J.S.D.); GreenLight Biosciences, Inc., 196 Boston Avenue, Suite 2400, Medford, Massachusetts 02155, USA (W.J.B.); Battelle Memorial Institute, 2987 Clairmont Road NE, Atlanta, Georgia 30329, USA (J.W.S.); School of Life Sciences, Tsinghua University, Beijing 100084, China (J.D.); Agilent Laboratories, 5301 Stevens Creek Boulevard, Mailstop 53L-IB, Santa Clara, California 95051, USA (D.L.L.); Bowdoin College, 5000 College Station, Brunswick, Maine 04011, USA (A.C.B.).

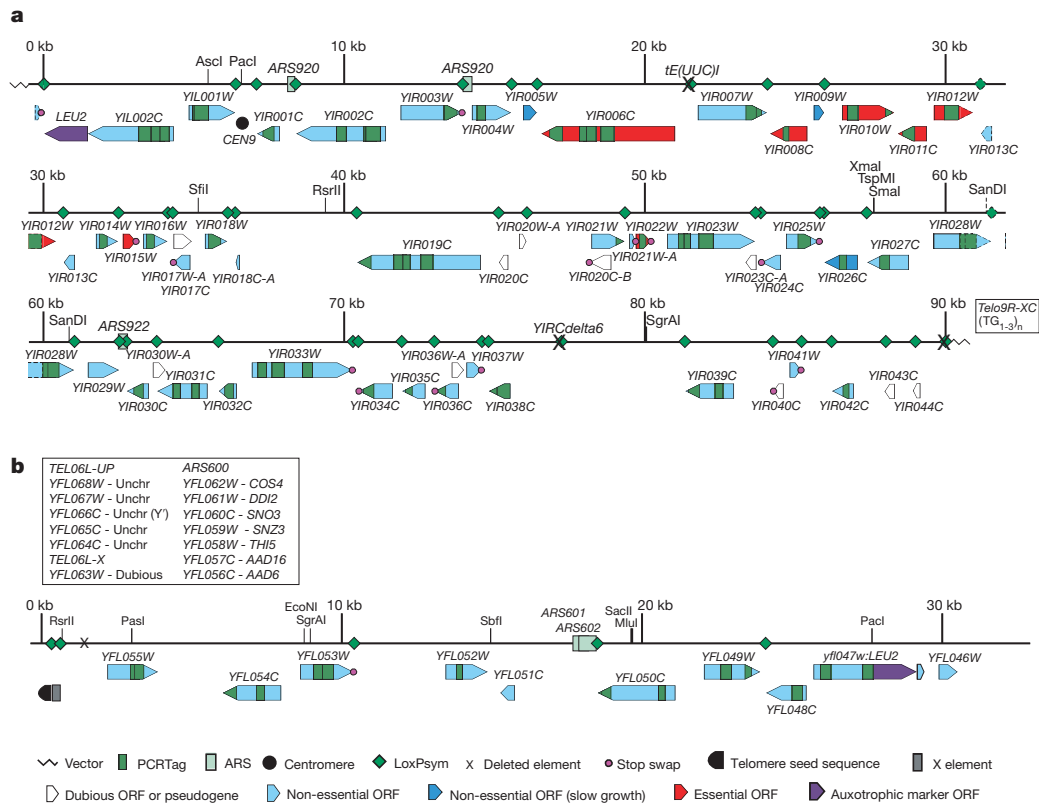


Figure 1 | Maps of synIXR and semi-synVII. Boxed text indicates elements deleted in the synthetic chromosomes. Vertical green bars inside ORFs indicate PCRTag amplicons; only sequences at the outside edges of these are recorded.

IXΔR. We observed more spore lethality than in control crosses, presumably owing to segregation of *synIXR* away from *IXΔR*; cells bearing only *synIXR* or only *IXΔR* would lack many essential genes and would not survive. PCRTag analysis of 14 *synIXR* candidate arm-swap strains revealed ten haplotypes with all synthetic PCRTags and no native PCRTags present (Fig. 2d and Supplementary Fig. 2). The remaining

ARS, autonomously replicating sequence. **a**, SynIXR. Vector is circular.
b, Semi-synVIL.

four strains carried BACs with patchworks of synthetic and native sequences indicative of meiotic gene-conversion events (Supplementary Fig. 2). Sanger sequencing and structural analyses (Supplementary Fig. 3, Supplementary Table 4 and Supplementary Text 3) of recovered synIXR BACs revealed that no mutations had occurred in the synthetic chromosome. Thus, the synthetic sequence is replicated faithfully.

Whereas synIXR was incorporated in a circular form, we used an alternate strategy to integrate the semi-synVIL chromosome fragment into native chromosome VI (Supplementary Fig. 4): a linear synthetic fragment marked with *LEU2* was transformed into a *YFL054C::kanMX* strain. Approximately 13% of transformants (75 of 586) had the $\text{Leu}^+\text{G418}^s$ phenotype expected for the desired integrant. PCRTag analysis showed that 10 of 12 such strains contained only synthetic PCRTags, as expected for full replacement (Supplementary Fig. 5).

The first design principle prioritizes a wild-type phenotype and a high level of fitness despite the incorporated modifications. SynIXR has a designed sequence alteration approximately every 500 bp, 2.64% of total sequence is altered, and it carries 43 loxPsym sites. To check for negative effects of modifications on fitness, we examined colony size and morphology under various conditions, and also performed transcript profiling. We inspected colony size and morphology of synIXR swap strains under six distinct growth conditions. It was impossible to

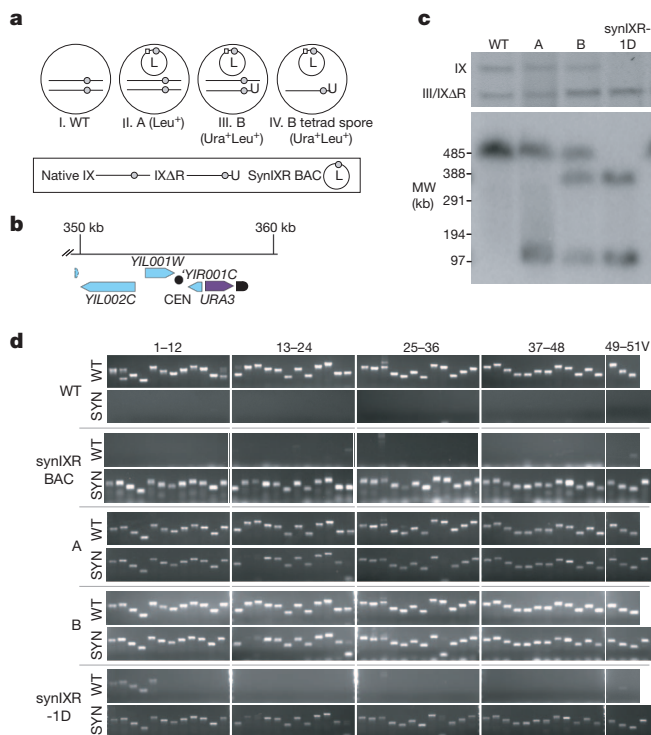


Figure 2 | Strain construction and verification. **a**, Generation of synIXR haploids. The synIXR BAC (L) was transformed into the wild-type strain BY4743 (WT, step I) to generate strain A (step II). One copy of native IXR in A was replaced with a *URA3*–telomere seed cassette (U), generating IXΔR in strain B (step III). B was sporulated to produce haploids (step IV). Circle, centromere; small square, *LEU2* gene. **b**, Structure of IXΔR. **c**, Electrophoretic karyotype (top panel) and Southern blot of NotI digest (bottom panel) of the wild-type, strain A, strain B and synIXR-1D genomes. Linearized synIXR migrates as a discrete band of ~100 kb. The probe (*YIL002C*) detects all isoforms of chromosome IX. *, native IXR; **, IXΔR. **d**, PCRTag analysis. SYN, synIXR BAC; V, vector amplicon.

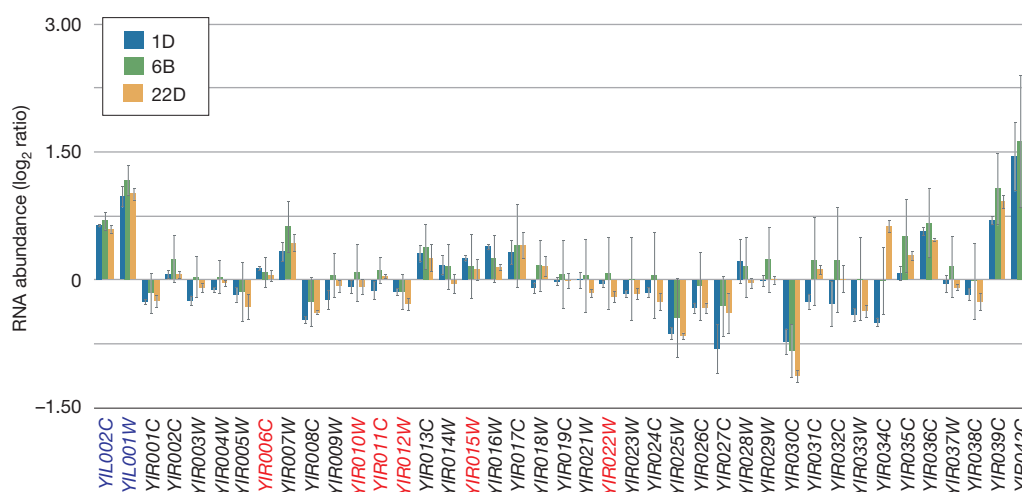


Figure 3 | Transcript profiling of wild-type and synIXR strains. Transcript profiling of synIXR-1D, -6B, and -22D. The log₂ ratio of RNA abundance relative to wild type (BY4741 or BY4742) is shown. YIL002C and YIL001W (blue) exist in two copies. Essential genes are labelled in red. Error bars, s.d.

distinguish swap strains from the wild type (BY4741) under these conditions, indicating that any fitness defect attributable to synIXR is modest; fitness tests on semi-synVIL gave similar results (Supplementary Fig. 6).

Synonymous substitutions, introduction of loxPsym sites or other changes might change gene expression. We performed transcript profiling on the swap strains synIXR-1D, synIXR-6B, and synIXR-22D (Supplementary Text 4); these studies revealed notable but predictable trends (Fig. 3). As expected, genes present in two copies (*YIL001W* and *YIL002C*, present on both synIXR and IXAR) were approximately doubled in transcript abundance. Most genes showed no substantial expression change, although a few showed modest decreases; however, the subtelomeric genes *YIR039C* and *YIR042C* showed increased expression. We speculate that in the circular synthetic chromosome, these are released from telomeric silencing, resulting in their overexpression. Overall, synIXR genes show relatively normal expression, indicating that loxPsym sites and PCRTags affect expression only minimally. Similarly, no substantial changes were observed by RNA blotting (Supplementary Fig. 7a). To detect possible compensatory transcriptome changes, we profiled transcripts genome-wide. Except for trivial differences attributable to slightly different configurations of selectable markers in the strains, there were no consistent, statistically significant differences outside IXR itself (Supplementary Fig. 7b). Thus, modifications present in synIXR and semi-synVIL do not produce major fitness effects or compensatory transcriptomic alterations.

A central feature of the synthetic yeast genome is the incorporated conditional genome instability system, SCRaMbLE. The design principles dictate that SCRaMbLE should be available for use on demand, yet should lie dormant until intentional Cre recombinase induction, at which point generation of genetic diversity is desirable. To complete the SCRaMbLE toolkit, we incorporated an engineered Cre recombinase fused to the murine oestrogen binding domain (EBD). This recently described Cre-EBD variant¹⁵ is oestradiol-inducible, has low basal activity and is controlled by the daughter-cell-specific promoter *SCW11* (Supplementary Fig. 8). The plasmid *pSCW11-Cre-EBD* should produce a pulse of recombinase activity once and only once in each cell's lifetime, and should depend on oestradiol exposure. The uninduced, integrated construct is well tolerated even in swap strains, which, with 43 loxPsym sites, are expected to be Cre-hypersensitive. Upon oestradiol addition, rearrangements were induced at the loxPsym sites and viability dropped by 100-fold in synIXR strains (Fig. 4a and Supplementary Fig. 9). This loss of viability probably results from loss of synIXR essential genes. In contrast, viability in semi-synVIL, which lacks essential genes, is not affected by Cre induction (Fig. 1b and Supplementary Fig. 9d).

Semi-synVIL contains just five loxPsym sites, including one immediately adjacent to the telomeric TG₁₋₃ repeats (Fig. 1b). This simple configuration allows comprehensive PCR-based mapping of rearrangements of four of the loxPsym sites in SCRaMbLEd strains. A SCRaMbLEd semi-synVIL population was analysed by PCR for most of the possible rearranged configurations, revealing a large variety of deletions and inversions (Fig. 4b); most predicted rearrangements were readily detected.

The symmetry of loxPsym sites allows alignment in two orientations, theoretically giving rise to deletions and inversions with equal frequency. SynIXR contains 43 loxPsym sites, allowing more than 3,600 potential pairwise interactions between synIXR loxPsym sites. We reasoned that SCRaMbLEd synIXR clones should display high phenotypic diversity. Indeed, SCRaMbLEd swap strains show more growth-rate heterogeneity than wild-type controls (Fig. 4c and Supplementary Fig. 10). These SCRaMbLEd clones show many different phenotypes (Supplementary Fig. 11 and Supplementary Text 5). In summary, SCRaMbLE is sufficient to generate substantial genetic heterogeneity and complex phenotypes.

To characterize the utility of SCRaMbLE further, we performed a mutagenesis study. SynIXR encodes both *MET28* and *LYS1*, genes required for biosynthesis of amino acids^{16,17}. Null mutants result in auxotrophy, and can be detected easily by replica-plating. We introduced episomal Cre-EBD (*pSCW11-Cre-EBD-URA3MX* cloned in a CEN plasmid) into strain C that was previously made *LYS2*⁺ (strain D, yJS587), and performed SCRaMbLE. We screened 20,242 colonies and 3% (604 of 20,242) were candidate *lys1* and/or *met28* auxotrophs. Of 360 candidates tested more rigorously, 295 (81.9%) were confirmed: we found 212 *Lys*⁻ auxotrophs (1.37%), 66 *Met*⁻ auxotrophs (0.43%) and, notably, 17 *Lys*⁻ *Met*⁻ double auxotrophs (0.11%). PCRTag profiles of 24 *Met*⁻ auxotrophs, 35 *Lys*⁻ auxotrophs and seven double auxotrophs (Fig. 4d) showed that all *Met*⁻ auxotrophs had deletions in the loxPsym-flanked segment containing *MET28* and *YAP5*, whereas all *Lys*⁻ auxotrophs had deletions in the loxPsym-flanked segment containing *LYS1*. The deletion profiles of many SCRaMbLEd auxotrophs were highly variable and more than one segment was often missing.

To confirm that the observed SCRaMbLE phenotypes resulted solely from deletions in synIXR, we recovered the synIXR chromosomes from two *Met*⁻ auxotrophs into *Escherichia coli*, and then introduced them to a clean genetic background. In both cases, the auxotrophic phenotype was associated with the presence of the SCRaMbLEd chromosomes (Supplementary Fig. 12 and Supplementary Text 6). Thus, the SCRaMbLE system is a highly effective method of mutagenesis, giving rise to mutants with different genetic backgrounds and generating a wide variety of double mutants.

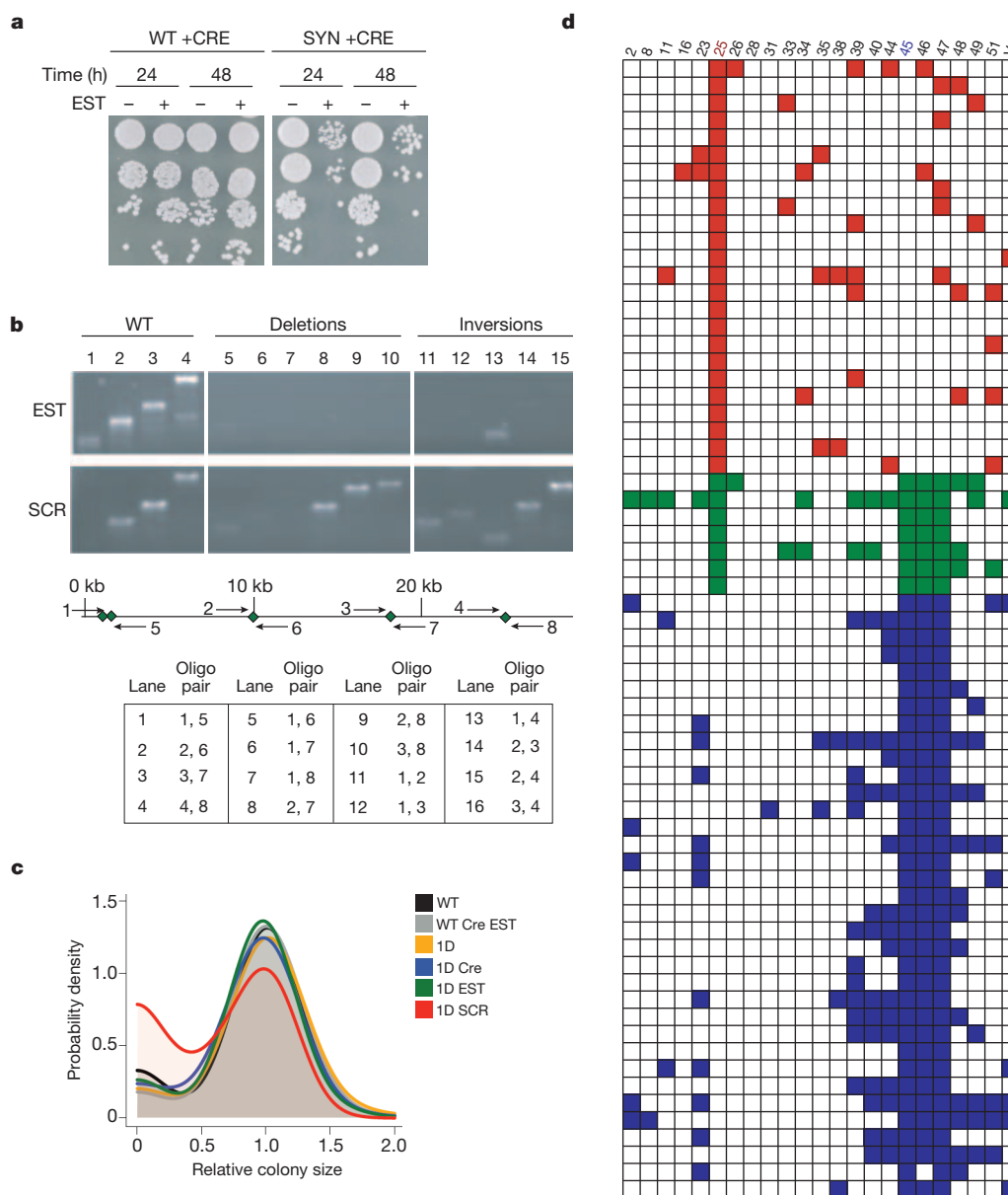


Figure 4 | SCRaMbLE rearranges genomes. a, Cre induction reduces the fitness of the synIXR strain (SYN) but not the wild type (WT; BY4741). EST, oestradiol; time, oestradiol exposure time. **b**, PCR analysis of semi-synVIL SCRaMbLE. The map shows primer positions. Amplicon 13 is spurious (wrong size). SCR, SCRaMbLE. **c**, Shifted colony-size distribution in SCRaMbLE survivors (wild type and the swap strain synIXR-1D). **d**, PCRTAG analysis of

Met⁻ (red), Lys⁻ (blue) and Met⁻ Lys⁻ (green) auxotrophs using PCRTags. PCRTAG pairs are numbered for each column (see Supplementary Table 2); MET28, pair 25; LYS1, pair 45. Each row represents one clone. Shaded boxes indicate presumed deletions. Panels a–c show strains with integrated Cre-EBD; d shows episomal Cre-EBD.

We have shown there does not seem to be any major theoretical impediment to extending the design strategy outlined here to the entire yeast genome, apart from the challenge of 12-megabase DNA synthesis. Whether or not fitness defects will accumulate as design and synthesis are scaled up remains to be seen; however, the overall high fitness of the swap strains described here validates the design strategy. Furthermore, the iterative, bottom-up approach will allow identification of potential ‘problem regions’ in synthetic sequences as synthesis moves forward. If a given swap experiment results in only transformants with reduced fitness (or if no transformants are obtainable), then the underlying defect can be mapped by introducing subsegments, facilitated by strategic placement of unique restriction sites throughout synthetic chromosome arms. Also, because a subset of transformants consist of patchworks of native and synthetic sequence (Supplementary Figs 2 and 5), analysis of such strains can be used to map phenotypic defects rapidly. The stability and sequence fidelity of

large circular chromosomes seen here and elsewhere^{5–7} bode well for the use of yeast as a host platform for synthetic biology.

SCRaMbLE may become a useful general strategy for analysing genome structure, content and function. One important feature of SCRaMbLE is its potential for customization: expression of different Cre-EBD variants from various promoters at distinct levels of inducer (oestradiol) should produce distinct SCRaMbLE dynamics. Use of weaker promoters than *pSCW11*, use of promoters expressed at different phases of the cell cycle, performing SCRaMbLE in diploids, and lowering the inducer concentration should all contribute to decreased lethality of SCRaMbLE strains, an important consideration as additional segments of the genome are replaced with synthetic counterparts and the proportion of essential genes that can be lost by SCRaMbLEing increases. As shown here, SCRaMbLE mutagenesis is efficient and generates mutants with a wide variety of different genetic backgrounds. It is possible that different combinations of gene deletions will give rise

BOX 1

Modifications in synthetic sequence

Elements removed

Retrotransposons: The *S. cerevisiae* genome contains both active retrotransposons and retrotransposon-derived sequences. These highly repetitive sequences are known to contribute to genome instability²². Because retrotransposons are presumed to be nonessential in yeast, we are eliminating these sequences from the synthetic genome.

Subtelomeric repeats: Two major types of subtelomeric repeats, Y' and X elements, reside in the genome. Y' elements are of unknown function, and are present at some, but not all, *S. cerevisiae* chromosome ends²³. In contrast, X elements are present in a single copy at all *S. cerevisiae* chromosome ends; they are more highly divergent, and function in telomeric silencing and possibly in chromosome segregation²³. To create a more streamlined genome, all Y' elements will be deleted from the synthetic genome; extant X elements will be replaced with the consensus core X-element sequence, as in semi-synVIL.

Introns: The yeast genome is estimated to contain approximately 285 introns. Based on a previous intron-deletion study²⁴ we do not anticipate that removal of introns will result in fitness defects; however, in some cases these introns house small non-coding RNAs (snoRNAs) that can be expressed ectopically in the synthetic yeast.

Elements relocated to extrachromosomal array

tRNA genes: tRNA genes (tDNAs) are highly redundant, with 275 nuclear tDNAs encoding only 42 tRNA species²⁵. In addition, these genes are known regions of genome instability^{8,9}. They will therefore be relocated to a dedicated chromosome to contain any instability resulting from their presence.

Elements replaced

TAG stop codons replaced by TAA: Removal of the TAG stop codon from the synthetic genome will allow future genetic code manipulation. The 'free' codon may be used to incorporate artificial amino acids^{11,12}; alternatively, the TAG codon may be placed in essential genes, and, exploiting an engineered orthogonal synthetase/tRNA pair, specify a non-genetically encoded amino acid, thereby providing a mechanism of reproductive isolation and an additional level of control over the synthetic yeast.

Individual synonymous codons: The synthetic genome is fabricated in fragments as small as 750 bp²⁶. Unique restriction sites are necessary within the synthetic fragment to facilitate construction of these building blocks into large contigs of up to 100 kb. Short stretches of fewer than four codons may therefore be synonymously recoded to introduce or eliminate restriction sites.

Strings of synonymous codons: Although several modifications exist between the native and synthetic genomes, the presence of a dedicated mechanism to distinguish between the two sequence types is invaluable. Short stretches of fewer than ten codons are therefore recoded to generate 'PCRTags', synonymous sequences used as the basis for PCR primer design to amplify selectively from wild-type or synthetic genomes.

Elements introduced

LoxPsym sites: Symmetrical loxP sites¹³ are inserted in the 3' UTR of all non-essential genes, as well as at synthetic landmarks. LoxPsym sites lack the directionality of canonical loxP sites, and can therefore align in two orientations. As a result, both inversions and deletions are predicted at equal probability. These loxPsym sites and an inducible Cre recombinase¹⁵ form the basis of the SCRaMbLE toolkit.

Elements not changed

Gene order: Gene order is preserved in the synthetic yeast to prevent incorporation of a non-permissible configuration in the design phase. Induction of SCRaMbLE results in changes in gene order and chromosome structure; all recovered SCRaMbLEd yeast have viable genome structures.

Noncoding regions: Except where noted, noncoding regions have not been modified. The yeast genome is well annotated; however, it is of paramount importance that the synthetic yeast be as fit as wild type until SCRaMbLE is induced. We therefore eschewed changes of noncoding regions to avoid disrupting unannotated critical elements. The few modifications that are made in noncoding sequence are kept to a minimum.

to a variety of subtly different phenotypes that can be mapped rapidly by PCRTag analysis; more extensive analysis by deep sequencing will reveal changes in genome structure and content. As the synthetic yeast genome grows, opportunities for genome rearrangement will increase exponentially. In principle, changes in chromosome number, ploidy, content and structure are all possible, increasing the utility of the SCRaMbLE system. For example, there may be many different routes to a minimal genome, and exploring all of them by a hit or miss predictive approach is impractical and unlikely to yield comprehensive results. Using SCRaMbLE, many independent routes of genome minimization can be explored at one time, under many environmental conditions, for instance by growing yeast cells long-term in serially transferred batch cultures, or in a chemostat or turbidostat under conditions in which Cre is minimally active. Such an approach may also lead to derivatives that are more fit than the parent, for example, by gene duplication events facilitated by the Cre-EBD/loxPsym system.

METHODS SUMMARY

DNA preparation. BAC DNA was prepared using the Qiagen plasmid midi kit or alkaline lysis¹⁸. The following protocol modifications were made: cells were diluted 1:100 from an overnight culture into 50 ml, grown in Luria broth with 50 µg ml⁻¹ carbenicillin, and grown at 30 °C for 14–16 h. Qiagen-purified DNA was treated with 60 µg ml⁻¹ proteinase K at 37 °C overnight, then extracted with phenol/chloroform. DNAs prepared without a column were phenol/chloroform extracted, and then treated with RNase immediately before use.

Yeast genomic DNA for use in PCRTag analysis was prepared by standard methods¹⁹. DNA preparation for recovery of the synIXR BAC into bacteria was as previously reported²⁰.

PCR conditions. PCRTags were amplified using Taq polymerase (New England Biolabs). Template concentrations were 1 ng µl⁻¹ for genomic DNA and 10 pg µl⁻¹ for purified BAC DNA. The following program was used: 94 °C 3 min; 30 cycles of 94 °C 30 s, 65 °C 30 s, 72 °C 30 s; 72 °C 3 min.

RNA analysis. Total RNA was isolated by hot acid phenol extraction. Microarray hybridization and data analysis were performed at the Johns Hopkins Microarray Core Facility (<http://www.microarray.jhmi.edu>). Dubious ORFs and pseudogenes were omitted from synIXR transcript analysis.

Pulsed-field gels. DNAs were prepared as described elsewhere²¹. The identity of the chromosomes was inferred from the known molecular karyotype of wild type (BY4743), and from lambda ladders run on the same gel.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 December 2010; accepted 28 June 2011.

Published online 14 September 2011.

1. Han, J. S. & Boeke, J. D. A highly active synthetic mammalian retrotransposon. *Nature* **429**, 314–318 (2004).
2. Richardson, S. M., Wheelan, S. J., Yarrington, R. M. & Boeke, J. D. GeneDesign: rapid, automated design of multikilobase synthetic genes. *Genome Res.* **16**, 550–556 (2006).
3. Chan, L. Y., Kosuri, S. & Endy, D. Refactoring bacteriophage T7. *Mol. Syst. Biol.* **1**, 2005.0018 (2005).

4. Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519 (2008).
5. Gibson, D. G. *et al.* One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl Acad. Sci. USA* **105**, 20404–20409 (2008).
6. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).
7. Lartigue, C. *et al.* Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* **325**, 1693–1696 (2009).
8. Ji, H. *et al.* Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. *Cell* **73**, 1007–1018 (1993).
9. Admire, A. *et al.* Cycles of chromosome instability are associated with a fragile site and are increased by defects in DNA replication and checkpoint controls in yeast. *Genes Dev.* **20**, 159–173 (2006).
10. Churcher, C. *et al.* The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IX. *Nature* **387**, 84–87 (1997).
11. Park, H. *et al.* Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science* **333**, 1151–1154 (2011).
12. Isaacs, F. J. *et al.* Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. *Science* **333**, 348–353 (2011).
13. Hoess, R. H., Wierzbicki, A. & Abremski, K. The role of the loxP spacer region in P1 site-specific recombination. *Nucleic Acids Res.* **14**, 2287–2300 (1986).
14. Vollrath, D., Davis, R. W., Connelly, C. & Hieter, P. Physical mapping of large DNA by chromosome fragmentation. *Proc. Natl Acad. Sci. USA* **85**, 6027–6031 (1988).
15. Lindstrom, D. L. & Gottschling, D. E. The mother enrichment program: a genetic system for facile replicative life span analysis in *Saccharomyces cerevisiae*. *Genetics* **183**, 413–422 (2009).
16. Kuras, L., Cherest, H., Surdin-Kerjan, Y. & Thomas, D. A heteromeric complex containing the centromere binding factor 1 and two basic leucine zipper factors, Met4 and Met28, mediates the transcription activation of yeast sulfur metabolism. *EMBO J.* **15**, 2519–2529 (1996).
17. Ogawa, H. & Fujioka, M. Purification and characterization of saccharopine dehydrogenase from baker's yeast. *J. Biol. Chem.* **253**, 3666–3670 (1978).
18. Sambrook, J. & Russell, D. W. Isolation of BAC DNA from small-scale cultures. *Cold Spring Harb. Protoc.* doi:10.1101/pdb.prot4006 (2006).
19. Hoffman, C. S. Preparation of yeast DNA. *Curr. Protoc. Mol. Biol.* Ch. 13, Unit 13.11 (2001).
20. Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. Ty elements transpose through an RNA intermediate. *Cell* **40**, 491–500 (1985).
21. Schwartz, D. C. & Cantor, C. R. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**, 67–75 (1984).
22. Lemoine, F. J., Degtyareva, N. P., Lobachev, K. & Petes, T. D. Chromosomal translocations in yeast induced by low levels of DNA polymerase: a model for chromosome fragile sites. *Cell* **120**, 587–598 (2005).
23. Louis, E. J. The chromosome ends of *Saccharomyces cerevisiae*. *Yeast* **11**, 1553–1573 (1995).
24. Parenteau, J. *et al.* Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Mol. Biol. Cell* **19**, 1932–1941 (2008).
25. Percudani, R., Pavesi, A. & Ottonello, S. Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **268**, 322–330 (1997).
26. Dymond, J. S. *et al.* Teaching synthetic biology, bioinformatics and engineering to undergraduates: the interdisciplinary build-a-genome course. *Genetics* **181**, 13–21 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Church for suggesting the global substitution of TAG codons with TAA codons, C. Connelly for sharing technical expertise and V. Huang for generating a sequence visualizer. We are grateful to B. Cormack, G. Seydoux and J. Nathans for offering helpful advice, to Y. Cai and J. Peccoud for suggesting methods to validate the sequence data, and to E. Louis for providing expert advice on telomeres. The work was supported by National Science Foundation grant MCB0718846 to J.D.B., J.S.B. and S.C.; by a grant from Microsoft to J.S.B. and J.D.B.; by Department of Energy Fellowship DE-FG02097ER25308 to S.M.R.; by National Institutes of Health grant AG023779 to D.E.G.; and by a fellowship from Fondation pour la Recherche Médicale to H.M.

Author Contributions J.S.D., S.M.R., S.C., J.S.B. and J.D.B. designed experiments. J.S.D., S.M.R., C.E.C., T.B., H.M., N.A., J.W.S., J.D. and A.C.B. performed experiments. W.J.B. built the synIXR chromosome. D.L.L. and D.E.G. generated the integrated CRE-EBD cassette. J.S.D., S.M.R., J.S.B. and J.D.B. analysed data and wrote the manuscript.

Author Information SynIXR and semi-synVIL sequences have been deposited to GenBank with the accession codes: synIXR, JN020955; semi-synVIL, JN020956. Microarray data have been submitted to Gene Expression Omnibus under accession number GSE31326. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.D.B. (jboeke@jhmi.edu).

METHODS

DNA preparation. BAC DNA was prepared using the Qiagen plasmid midi kit or alkaline lysis¹⁸. The following protocol modifications were made: cells were diluted 1:100 from an overnight culture into 50 ml, grown in Luria broth with 50 µg ml⁻¹ carbenicillin, and grown at 30 °C for 14–16 h. Qiagen-purified DNA was treated with 60 µg ml⁻¹ proteinase K at 37 °C overnight, then extracted with phenol/chloroform. DNAs prepared without a column were phenol/chloroform extracted, and then treated with RNase immediately before use.

Yeast genomic DNA for use in PCRTAG analysis was prepared by standard methods¹⁹. DNA preparation for recovery of the synIXR BAC into bacteria was as previously reported²⁰.

PCR conditions. PCRTags were amplified using Taq polymerase (New England Biolabs). Template concentrations were 1 ng µl⁻¹ for genomic DNA and 10 pg µl⁻¹ for purified BAC DNA. The following program was used: 94 °C 3 min; 30 cycles of 94 °C 30 s, 65 °C 30 s, 72 °C 30 s; 72 °C 3 min.

RNA analysis. Total RNA was isolated by hot acid phenol extraction. Microarray hybridization and data analysis were performed at the Johns Hopkins Microarray Core Facility (<http://www.microarray.jhmi.edu>). Dubious ORFs and pseudogenes were omitted from synIXR transcript analysis.

Pulsed-field gels. DNAs were prepared as described elsewhere²¹. The identity of the chromosomes was inferred from the known molecular karyotype of wild type (BY4743), and from lambda ladders run on the same gel.

Yeast strains, transformation and tetrad analysis. Strains ABY7 and ABY8 were derived from strain BY4743; ABY7 (*MATa*) and ABY8 (*MATα*) otherwise share the genotype *his3Δ1 leu2Δ0 ura3Δ0 lys2Δ0 met15Δ0 yil001::URA3 yir039::kanMX*. All strain genotypes are listed in Supplementary Table 8.

BY4743 spheroplasts were transformed with synIXR. The strain *YFL054C::kanMX* was transformed with synVII restriction fragments by standard lithium acetate transformation.

The synIXR-1D strain and others were backcrossed to strains ABY7 and ABY8; the resultant diploids were sporulated and genotyped to identify synIXR segregants. **Phenotypic screening.** Single colonies were picked into 96-well plates and grown for 48 h in yeast peptone dextrose (YPD) at 30 °C. (SCRaMbLE strains were grown for 72 h in YPD at 30 °C, diluted 1:10 and grown for 4 h before plating.) Tenfold dilutions were spotted on various types of agar medium and selective conditions in OmniTrays (NUNC), as previously described²⁷. Most cells were grown for 72 h (except those grown on yeast extract/peptone/glycerol/ethanol (YPGE) plates, which were grown for 108 h), then scored for growth and photographed.

Yeast growth and media. Unless otherwise indicated, all experiments were performed at 30 °C. YPGE was supplemented with 2% ethanol and 2% glycerol. Concentrations of drugs were as follows: hydroxyurea, 0.2 M; methylmethane

sulphonate, 0.05%; 6-azauracil, 100 µg ml⁻¹; benomyl, 15 µg ml⁻¹; hydrogen peroxide, 1 mM; cycloheximide, 10 µg ml⁻¹. Resistance to cycloheximide and hydrogen peroxide was assayed by growing cells in treated medium for 2 h, then plating on YPD. Other phenotypes were assayed by growing cells to mid-log phase in rich media, then spotting tenfold dilutions on selective media.

Colony size measurements. Cells were plated at various dilutions so that similar numbers of colonies were observed on control and experimental (oestradiol-treated) plates. Colony size was measured using ImageJ software²⁸, and normalized against the total number of colonies on each plate. Sample sizes for data presented in Fig. 4c are as follows: wild-type, *n* = 488 colonies; wild-type + Cre + oestradiol, *n* = 486; 1D, *n* = 395; 1D + Cre, *n* = 251; 1D + oestradiol, *n* = 416; 1D + Cre + oestradiol, *n* = 394.

SynIXR BAC sequence analysis. The original synIXR BAC was sequenced by the manufacturer, Codon Devices²⁹. SynIXR BACs were recovered into bacteria and sequenced by Agencourt (Beckman Coulter Genomics), using sequencing primers listed in Supplementary Table 5. Repetitive sequences, including the highly internally repetitive *MUC1* open reading frame, were PCR-amplified before sequencing when necessary.

Pulsed-field gels. Samples were run on a 1.0% agarose gel in ×0.5 TBE (pH 8.0) for 20 h at 14 °C on a clamped homogenous electric field (CHEF) gel apparatus. The voltage was 3.5 V cm⁻¹, at an angle of 120° and a switch time of 60–120 s, ramped over 20 h.

NotI (Promega) digests were performed on whole chromosomes embedded in agarose plugs. Agarose plugs were removed from the 0.5 M EDTA storage buffer, washed with 0.05 M EDTA for 1 h at room temperature (~23 °C), and then washed with ×0.1 restriction enzyme buffer, followed by ×1 buffer, under the same conditions.

Probe preparation for northern and Southern blots. Probes were prepared using the Prime-It II kit (Stratagene) and hybridized using UltraHyb hybridization solution (Ambion) according to the manufacturer's instructions.

SCRaMbLE. Cre activity was induced by exposure to 1 µM β-oestradiol (Sigma-Aldrich) in rich media for either 48 h (integrated Cre) or 4 h (episomal Cre), except where indicated otherwise. PCRTAG analysis of Met⁻ and Lys⁻ auxotrophs was performed with a non-redundant array, using one primer pair per loxPsym-flanked segment.

27. Hampsey, M. A review of phenotypes in *Saccharomyces cerevisiae*. *Yeast* **13**, 1099–1133 (1997).
28. Abramoff, M. D., Magelhaes, P. J. & Ram, S. J. Image processing with ImageJ. *Biophotonics Int.* **11**, 36–42 (2004).
29. Blake, W. J. *et al.* Pairwise selection assembly for sequence-independent construction of long-length DNA. *Nucleic Acids Res.* **38**, 2594–2602 (2010).

Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk

The International Consortium for Blood Pressure Genome-Wide Association Studies

Blood pressure is a heritable trait¹ influenced by several biological pathways and responsive to environmental stimuli. Over one billion people worldwide have hypertension (≥ 140 mm Hg systolic blood pressure or ≥ 90 mm Hg diastolic blood pressure)². Even small increments in blood pressure are associated with an increased risk of cardiovascular events³. This genome-wide association study of systolic and diastolic blood pressure, which used a multi-stage design in 200,000 individuals of European descent, identified sixteen novel loci: six of these loci contain genes previously known or suspected to regulate blood pressure (*GUCY1A3–GUCY1B3*, *NPR3–C5orf23*, *ADM*, *FURIN–FES*, *GOSR2*, *GNAS–EDN3*); the other ten provide new clues to blood pressure physiology. A genetic risk score based on 29 genome-wide significant variants was associated with hypertension, left ventricular wall thickness, stroke and coronary artery disease, but not kidney disease or kidney function. We also observed associations with blood pressure in East Asian, South Asian and African ancestry individuals. Our findings provide new insights into the genetics and biology of blood pressure, and suggest potential novel therapeutic pathways for cardiovascular disease prevention.

Genetic approaches have advanced the understanding of biological pathways underlying inter-individual variation in blood pressure. For example, studies of rare Mendelian blood pressure disorders have identified multiple defects in renal sodium handling pathways⁴. More recently two genome-wide association studies (GWAS), each of >25,000 individuals of European ancestry, identified 13 loci associated with systolic blood pressure (SBP), diastolic blood pressure (DBP) and hypertension^{5,6}. We now report results of a new meta-analysis of GWAS data that includes staged follow-up genotyping to identify additional blood pressure loci.

Primary analyses evaluated associations between 2.5 million genotyped or imputed single nucleotide polymorphisms (SNPs) and SBP and DBP in 69,395 individuals of European ancestry from 29 studies (Supplementary Materials sections 1–3 and Supplementary Tables 1 and 2). Following GWAS meta-analysis, we conducted a three-stage validation experiment that made efficient use of available genotyping resources, to follow up top signals in up to 133,661 additional individuals of European descent (Supplementary Fig. 1 and Supplementary Materials section 4). Twenty-nine independent SNPs at 28 loci were significantly associated with SBP, DBP, or both in the meta-analysis combining discovery and follow-up data (Fig. 1, Table 1, Supplementary Figs 2, 3 and Supplementary Tables 3–5). All 29 SNPs attained association $P < 5 \times 10^{-9}$, an order of magnitude beyond the standard genome-wide significance level for a single-stage experiment (Table 1).

Sixteen of these 29 associations were novel (Table 1). Two associations were near the *FURIN* and *GOSR2* genes; prior targeted analyses of variants in these genes suggested they may be blood pressure loci^{7,8}. At the *CACNB2* locus we validated association for a previously reported⁶ SNP, rs4373814, and detected a novel independent association for rs1813353 (pairwise $r^2 = 0.015$ in HapMap CEU). Of our 13 previously reported associations^{5,6}, only the association at *PLCD3*

was not supported by the current results (Supplementary Table 4). Some of the associations are in or near genes involved in pathways known to influence blood pressure (*NPR3*, *GUCY1A3–GUCY1B3*, *ADM*, *GNAS–EDN3*, *NPPA–NPPB* and *CYP17A1*; Supplementary Fig. 4). Twenty-two of the 28 loci did not contain genes that were a priori strong biological candidates.

As expected from prior blood pressure GWAS results, the effects of the novel variants on SBP and DBP were small (Fig. 1 and Table 1). For all variants, the observed directions of effects were concordant for SBP, DBP and hypertension (Fig. 1, Table 1 and Supplementary Fig. 3). Among the genes at the genome-wide significant loci, only *CYP17A1*, previously implicated in Mendelian congenital adrenal hyperplasia and hypertension, is known to harbour rare variants that have large effects on blood pressure⁹.

We performed several analyses to identify potential causal alleles and mechanisms. First, we looked up the 29 genome-wide significant index SNPs and their close proxies ($r^2 > 0.8$) among *cis*-acting expression SNP (eSNP) results from multiple tissues (Supplementary Materials section 5). For 13/29 index SNPs, we found an association between nearby eSNP variants and the expression levels of at least one gene transcript ($10^{-4} > P > 10^{-51}$; Supplementary Table 6). In five cases, the index blood pressure SNP and the best eSNP from a genome-wide survey were identical, highlighting potential mediators of the SNP–blood pressure associations.

Second, because changes in protein sequence are a priori strong functional candidates, we sought non-synonymous coding SNPs that were in high linkage disequilibrium ($r^2 > 0.8$) with the 29 index SNPs. We identified such SNPs at eight loci (Table 1, Supplementary Materials section 6 and Supplementary Table 7). In addition we performed analyses testing for differences in genetic effect according to body mass index (BMI) or sex, and analyses of copy number variants, pathway enrichment and metabolomic data, but we did not find any statistically significant results (Supplementary Materials sections 7–9 and Supplementary Tables 8–10).

We evaluated whether the blood pressure variants we identified in individuals of European ancestry were associated with blood pressure in individuals of East Asian ($N = 29,719$), South Asian ($N = 23,977$) and African ($N = 19,775$) ancestries (Table 1 and Supplementary Tables 11–13). We found significant associations in individuals of East Asian ancestry for SNPs at nine loci and in individuals of South Asian ancestry for SNPs at six loci; some have been reported previously (Supplementary Tables 12 and 15). The lack of significant association for individual SNPs may reflect small sample sizes, differences in allele frequencies or linkage disequilibrium patterns, imprecise imputation for some ancestries using existing reference samples, or a genuinely different underlying genetic architecture. Because of limited power to detect effects of individual variants in the smaller non-European samples, we created genetic risk scores for SBP and DBP incorporating all 29 blood pressure variants weighted according to effect sizes observed in the European samples. In each non-European ancestry group, risk scores were strongly associated with SBP ($P = 1.1 \times 10^{-40}$ in East Asian, $P = 2.9 \times 10^{-13}$ in South Asian, $P = 9.8 \times 10^{-4}$ in African

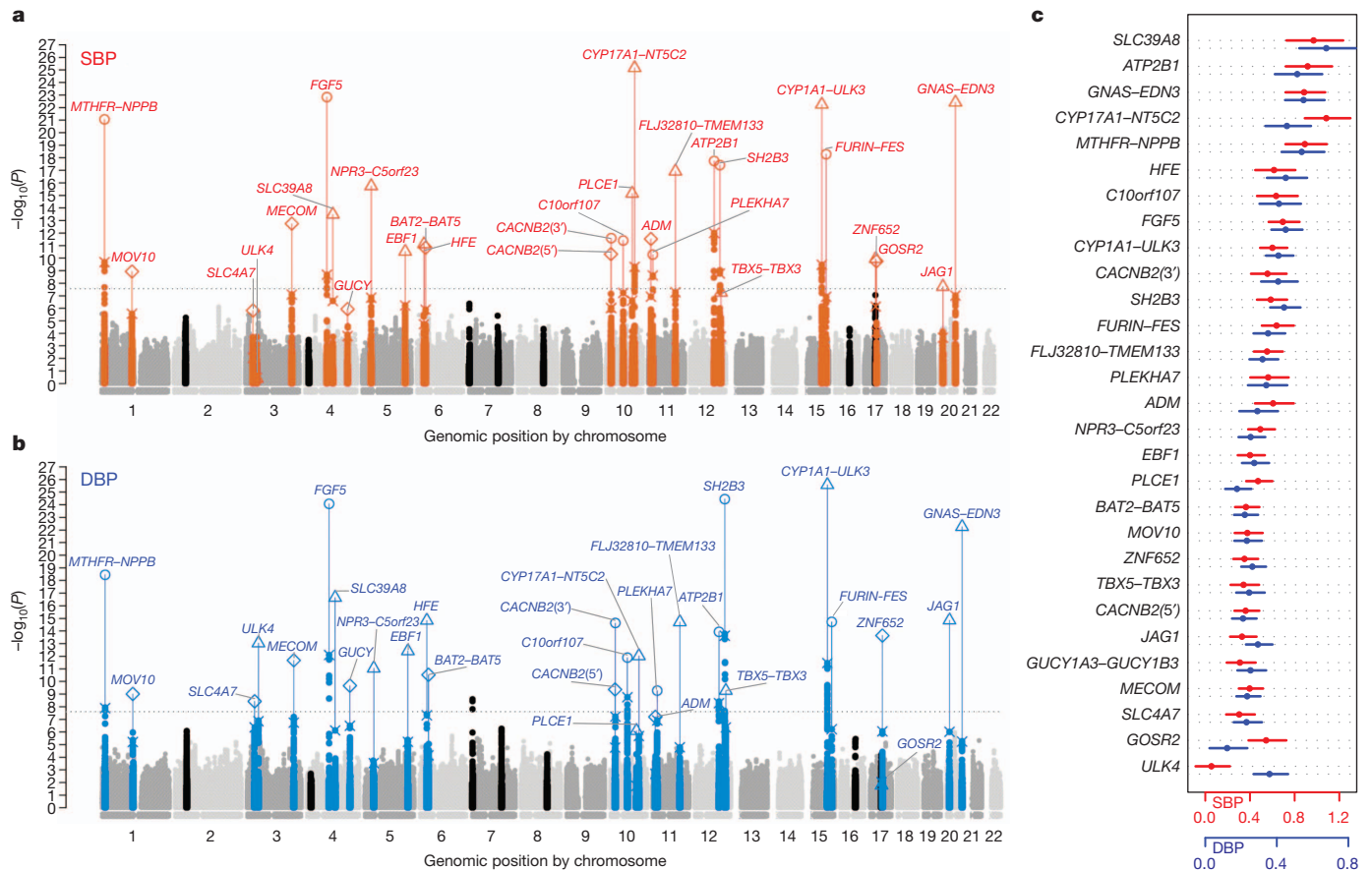


Figure 1 | Genome-wide $-\log_{10} P$ -value plots and effects for significant loci. a, b, Genome-wide $-\log_{10} P$ -value plots are shown for SBP (a) and DBP (b). SNPs within loci reaching genome-wide significance are labelled in red for SBP and blue for DBP (± 2.5 Mb of lowest P value) and lowest P values in the initial genome-wide analysis as well as the results of analysis including validation data are labelled separately. The lowest P values in the initial GWAS are denoted with a X. The range of different sample sizes in the final meta-

ancestry individuals) and DBP ($P = 2.9 \times 10^{-48}$, $P = 9.5 \times 10^{-15}$ and $P = 5.3 \times 10^{-5}$, respectively; Supplementary Table 13).

We also created a genetic risk score to assess association of the variants in aggregate with hypertension and with clinical measures of hypertensive complications including left ventricular mass, left ventricular wall thickness, incident heart failure, incident and prevalent stroke, prevalent coronary artery disease (CAD), kidney disease and measures of kidney function, using results from other GWAS consortia (Table 2, Supplementary Materials sections 10, 11 and Supplementary Table 14). The risk score was weighted using the average of SBP and DBP effects for the 29 SNPs. In an independent sample of 23,294 women¹⁰, an increase of one standard deviation in the genetic risk score was associated with a 23% increase in the odds of hypertension (95% confidence interval 19–28%; Table 2 and Supplementary Table 14). Among individuals in the top decile of the risk score, the prevalence of hypertension was 29% compared with 16% in the bottom decile (odds ratio 2.09, 95% confidence interval 1.86–2.36). Similar results were observed in an independent hypertension case-control sample (Table 2). In our study, individuals in the top compared to bottom quintiles of genetic risk score differed by 4.6 mm Hg SBP and 3.0 mm Hg DBP, differences that approach population-averaged blood pressure treatment effects for a single antihypertensive agent¹¹. Epidemiological data have shown that differences in SBP and DBP of this magnitude, across the population range of blood pressure, are associated with an increase in cardiovascular disease risk³. Consistent with this and in line with findings from randomized trials

analysis including the validation data are indicated as: circle (96,000–140,000), triangle ($>140,000$ –180,000) and diamond ($>180,000$ –220,000). SNPs near unconfirmed loci are in black. The horizontal dotted line is $P = 2.5 \times 10^{-8}$. GUCY denotes GUCY1A3–GUCY1B3. c, Effect size estimates and 95% confidence bars per blood-pressure-increasing allele of the 29 significant variants for SBP (red) and DBP (blue). Effect sizes are expressed in mm Hg per allele.

of blood-pressure-lowering medication in hypertensive patients^{12,13}, the genetic risk score was positively associated with left ventricular wall thickness ($P = 6.0 \times 10^{-6}$), occurrence of stroke ($P = 3.3 \times 10^{-5}$) and CAD ($P = 8.1 \times 10^{-29}$). The same genetic risk score was not, however, significantly associated with chronic kidney disease or measures of kidney function, even though these renal outcomes were available in a similar sample size as for the other outcomes (Table 2). The absence of association with kidney phenotypes could be explained by a weaker causal relationship between blood pressure and kidney phenotypes than with CAD and stroke. This finding is consistent with the mismatch between observational data that show a positive association of blood pressure with kidney disease, and clinical trial data that show inconsistent evidence of a benefit from blood pressure lowering on kidney disease prevention in patients with hypertension¹⁴. Thus, several lines of evidence converge to indicate that blood pressure elevation may in part be a consequence rather than a cause of sub-clinical kidney disease.

Our discovery meta-analysis (Supplementary Fig. 2) suggests an excess of modestly significant ($10^{-5} < P < 10^{-2}$) associations probably arising from common blood pressure variants of small effect. By dividing our principal GWAS data set into non-overlapping discovery ($N \approx 56,000$) and validation ($N \approx 14,000$) subsets, we found robust evidence for the existence of such undetected common variants (Supplementary Fig. 5 and Supplementary Materials section 12). We estimate¹⁵ that there are 116 (95% confidence interval 57–174) independent blood pressure variants with effect sizes similar to those

Table 1 | Summary association results for 29 blood pressure SNPs

Locus	Index SNP	Chr	Position	CA/ NCA	CAF	nsSNP	eSNP	SBP			DBP			HTN	
								Beta	P value	Effect in EA/SA/A	Beta	P value	Effect in EA/SA/A	Beta	P value
<i>MOV10</i>	rs2932538	1	113,018,066	G/A	0.75	Y(p)	Y(p)	0.388	1.2×10^{-9}	+/+/-	0.240	9.9×10^{-10}	+/+*/-	0.049	2.9×10^{-7}
<i>SLC4A7</i>	rs13082711	3	27,512,913	T/C	0.78	Y(p)	Y(p)	-0.315	1.5×10^{-6}	-/-/+	-0.238	3.8×10^{-9}	-/-/+	-0.035	3.6×10^{-4}
<i>MECOM</i>	rs419076	3	170,583,580	T/C	0.47	-	-	0.409	1.8×10^{-13}	+/+/+	0.241	2.1×10^{-12}	+/+/-	0.031	3.1×10^{-4}
<i>SLC39A8</i>	rs13107325	4	103,407,732	T/C	0.05	Y	Y(+)	-0.981	3.3×10^{-14}	?/+/?	-0.684	2.3×10^{-17}	?/+/?	-0.105	4.9×10^{-7}
<i>GUCY1A3- GUCY1B3</i>	rs13139571	4	156,864,963	C/A	0.76	-	-	0.321	1.2×10^{-6}	+/-/+	0.260	2.2×10^{-10}	+/-/+	0.042	2.5×10^{-5}
<i>NPR3- C5orf23</i>	rs1173771	5	32,850,785	G/A	0.60	-	-	0.504	1.8×10^{-16}	+*/+/+	0.261	9.1×10^{-12}	+*/+/-	0.062	3.2×10^{-10}
<i>EBF1</i>	rs11953630	5	157,777,980	T/C	0.37	-	-	-0.412	3.0×10^{-11}	+/+/+	-0.281	3.8×10^{-13}	+/+/+	-0.052	1.7×10^{-7}
<i>HFE</i>	rs1799945	6	26,199,158	G/C	0.14	Y	-	0.627	7.7×10^{-12}	+/+/-	0.457	1.5×10^{-15}	+/+/-	0.095	1.8×10^{-10}
<i>BAT2-BAT5</i>	rs805303	6	31,724,345	G/A	0.61	Y(p)	Y(+)	0.376	1.5×10^{-11}	-/-/?	0.228	3.0×10^{-11}	-/-/+	0.054	1.1×10^{-10}
<i>CACNB2(5')</i>	rs4373814	10	18,459,978	G/C	0.55	-	-	-0.373	4.8×10^{-11}	+*/+/-	-0.218	4.4×10^{-10}	+*/+/-	-0.046	8.5×10^{-8}
<i>PLCE1</i>	rs932764	10	95,885,930	G/A	0.44	-	-	0.484	7.1×10^{-16}	+/+/-	0.185	8.1×10^{-7}	+/+/-	0.055	9.4×10^{-9}
<i>ADM</i>	rs7129220	11	10,307,114	G/A	0.89	-	-	-0.619	3.0×10^{-12}	?/-/+	-0.299	6.4×10^{-8}	?/-/+	-0.044	1.1×10^{-3}
<i>FLJ32810- TMEM133</i>	rs633185	11	100,098,748	G/C	0.28	-	-	-0.565	1.2×10^{-17}	+*/+/+	-0.328	2.0×10^{-15}	+*/+/-	-0.070	5.4×10^{-11}
<i>FURIN-FES</i>	rs2521501	15	89,238,392	T/A	0.31	-	Y(-)	0.650	5.2×10^{-19}	+*/+/+	0.359	1.9×10^{-15}	+*/+/+	0.059	7.0×10^{-7}
<i>GOSR2</i>	rs17608766	17	42,368,270	T/C	0.86	-	Y(+)	-0.556	1.1×10^{-10}	+*/+/-	-0.129	0.017	+*/+/-	-0.025	0.08
<i>JAG1</i>	rs1327235	20	10,917,030	G/A	0.46	-	-	0.340	1.9×10^{-8}	+*/+/+	0.302	1.4×10^{-15}	+*/+/+	0.034	4.6×10^{-4}
<i>GNAS-EDN3</i>	rs6015450	20	57,184,512	G/A	0.12	Y(p)	-	0.896	3.9×10^{-23}	?/+/?	0.557	5.6×10^{-23}	?/+/?	0.110	4.2×10^{-14}
<i>MTHFR- NPPB</i>	rs17367504	1	11,785,365	G/A	0.15	-	Y(-/r)	-0.903	8.7×10^{-22}	+/+/+	-0.547	3.5×10^{-19}	+/+/+	-0.103	2.3×10^{-10}
<i>ULK4</i>	rs3774372	3	41,852,418	T/C	0.83	Y	Y(r/p)	-0.067	0.39	-/-/+	-0.367	9.0×10^{-14}	+/+/+	-0.017	0.18
<i>FGF5</i>	rs1458038	4	81,383,747	T/C	0.29	-	-	0.706	1.5×10^{-23}	+*/+/+	0.457	8.5×10^{-25}	+*/+/+	0.072	1.9×10^{-7}
<i>CACNB2(3')</i>	rs1813353	10	18,747,454	T/C	0.68	-	-	0.569	2.6×10^{-12}	+/+/+	0.415	2.3×10^{-15}	+/+/+	0.078	6.2×10^{-10}
<i>C10orf107</i>	rs4590817	10	63,137,559	G/C	0.84	-	Y(r)	0.646	4.0×10^{-12}	-/+/-	0.419	1.3×10^{-12}	-/-/-	0.096	9.8×10^{-9}
<i>CYP17A1- NT5C2</i>	rs11191548	10	104,836,168	T/C	0.91	-	Y(-)	1.095	6.9×10^{-26}	+*/+/+	0.464	9.4×10^{-13}	+*/+/+	0.097	1.4×10^{-5}
<i>PLEKHA7</i>	rs381815	11	16,858,844	T/C	0.26	-	-	0.575	5.3×10^{-11}	+*/+/+	0.348	5.3×10^{-10}	+*/+/-	0.062	3.4×10^{-6}
<i>ATP2B1</i>	rs17249754	12	88,584,717	G/A	0.84	-	-	0.928	1.8×10^{-18}	+*/+*/-	0.522	1.2×10^{-14}	+*/+*/-	0.126	1.1×10^{-14}
<i>SH2B3</i>	rs3184504	12	110,368,991	T/C	0.47	Y	Y(+)	0.598	3.8×10^{-18}	-/-/+	0.448	3.6×10^{-25}	-/-/+	0.056	2.6×10^{-6}
<i>TBX5-TBX3</i>	rs10850411	12	113,872,179	T/C	0.7	-	-	0.354	5.4×10^{-8}	-/+/-	0.253	5.4×10^{-10}	-/-/-	0.045	5.2×10^{-6}
<i>CYP1A1- ULK3</i>	rs1378942	15	72,864,420	C/A	0.35	-	Y(+)	0.613	5.7×10^{-23}	+*/+/+	0.416	2.7×10^{-26}	+*/+/-	0.073	1.0×10^{-8}
<i>ZNF652</i>	rs12940887	17	44,757,806	T/C	0.38	-	Y(-)	0.362	1.8×10^{-10}	+/-/+	0.27	2.3×10^{-14}	+/-/+	0.046	1.2×10^{-7}

Summary association statistics, based on combined discovery and follow-up data, for 29 independent SNPs in individuals of European ancestry are shown. New genome-wide significant findings (17 SNPs) are presented in the top half of the table, data on 12 previously published signals are presented in the lower half. Y indicates that the blood pressure index SNP is a non-synonymous (ns)SNP, Y(p) indicates a proxy SNP is a nsSNP. Y(+) indicates that the blood pressure index SNP is the strongest known eSNP for a transcript; Y(-) indicates that the blood pressure index SNP is an eSNP but not the strongest known eSNP for any transcript. Y(r) indicates that the blood pressure index SNP is the strongest known eSNP in a targeted real-time PCR experiment. Y(p) indicates that a proxy SNP ($r^2 > 0.8$) to a blood pressure SNP is an eSNP but not the strongest known eSNP. Observed effect directions in East Asian (EA), South Asian (SA) and African (A) ancestry individuals are coded + or - if concordant or discordant with directions in European ancestry results. Effect size estimates (beta) correspond to mm Hg per coded allele for SBP and DBP and ln(odds) per coded allele for hypertension (HTN). CA, coded allele; CAF, coded allele frequency; NCA, non-coded allele. ? denotes missing data. Genomic positions use NCBI Build 36 coordinates.

* Significant, controlling the FDR at 5% over 58 tests per ancestry (Supplementary Tables 5 and 12).

reported here, which collectively can explain ~2.2% of the phenotypic variance for SBP and DBP, compared with 0.9% explained by the 29 associations discovered thus far (Supplementary Fig. 6 and Supplementary Materials section 13).

Most of the 28 blood pressure loci harbour multiple genes (Supplementary Table 15 and Supplementary Fig. 4), and although substantial research is required to identify the specific genes and variants responsible for these associations, several loci contain highly plausible biological candidates. The *NPPA* and *NPPB* genes at the *MTHFR-NPPB* locus encode precursors for atrial- and B-type natriuretic peptides (ANP, BNP), and previous work has identified SNPs—modestly correlated with our index SNP at this locus—which are associated with plasma ANP, BNP and blood pressure¹⁶. We found the index SNP at this locus was associated with opposite effects on blood pressure and on ANP/BNP levels, consistent with a model in which the variants act through increased ANP/BNP production to lower blood pressure¹⁶ (Supplementary Materials section 14).

Two other loci identified in the current study harbour genes involved in natriuretic peptide and related nitric oxide signalling pathways^{17,18}, both of which act to regulate cyclic guanosine monophosphate. The first locus contains *NPR3*, which encodes the natriuretic peptide clearance receptor (NPR-C). *NPR3* knockout mice exhibit reduced clearance of circulating natriuretic peptides and lower blood pressure¹⁹. The second locus includes *GUCY1A3* and *GUCY1B3*, encoding the α and β subunits of soluble guanylate cyclase; knockout of either gene in murine models results in hypertension²⁰.

Another locus contains *ADM*—encoding adrenomedullin—which has natriuretic, vasodilatory and blood-pressure-lowering properties²¹. At the *GNAS-EDN3* locus, *ZNF831* is closest to the index SNP, but *GNAS* and *EDN3* are two nearby compelling biological candidates (Supplementary Fig. 4 and Supplementary Table 15).

We identified two loci with plausible connections to blood pressure via genes implicated in renal physiology or kidney disease. At the first locus, *SLC4A7* is an electro-neutral sodium bicarbonate co-transporter expressed in the nephron and in vascular smooth muscle²². At the second locus, *PLCE1* (phospholipase-C-epsilon-1 isoform) is important for normal podocyte development in the glomerulus; sequence variation in *PLCE1* has been implicated in familial nephrotic syndromes and end-stage kidney disease²³.

Missense variants in two genes involved in metal ion transport were associated with blood pressure in our study. The first encodes a His/Asp change at amino acid 63 (H63D) in *HFE* and is a low-penetrance allele for hereditary hemochromatosis²⁴. The second is an Ala/Thr polymorphism located in exon 7 of *SLC39A8*, which encodes a zinc transporter that also transports cadmium and manganese²⁵. The same allele of *SLC39A8* associated with blood pressure in our study has recently been associated with high-density lipoprotein cholesterol levels²⁶ and BMI²⁷ (Supplementary Table 15).

We have shown that 29 independent genetic variants influence blood pressure in people of European ancestry. The variants reside in 28 loci, 16 of which were novel, and we confirmed association of several of them in individuals of non-European ancestry. A risk score

Table 2 | Genetic risk score and cardiovascular outcome association results

Phenotype	Source	Effect		s.e.	P value	No. SNPs	Contrast top versus bottom			N case/control or total
		(per s.d. of genetic risk score)					Quintiles	Deciles		
Blood pressure phenotypes										
SBP (mm Hg)	WGHS	1.645	0.098	(a)	6.5×10^{-63}	29	4.61	5.77	(a)	23,294
DBP (mm Hg)	WGHS	1.057	0.067	(a)	8.4×10^{-57}	29	2.96	3.71	(a)	23,294
Prevalent hypertension	WGHS	0.211	0.018	(b)	3.1×10^{-33}	29	1.80	2.09	(b)	5,018/18,276
Prevalent hypertension	BRIGHT	0.287	0.031	(b)	7.7×10^{-21}	29	2.23	2.74	(b)	2,406/1,990
Dichotomous endpoints										
Incident heart failure	CHARGE-HF	0.035	0.021	(c)	0.10	29	1.10	1.13	(c)	2,526/18,400
Incident stroke	NEURO-CHARGE	0.103	0.028	(c)	0.0002	28	1.34	1.44	(c)	1,544/18,058
Prevalent stroke	SCG	0.075	0.037	(b)	0.05	29	1.23	1.30	(b)	1,473/1,482
Stroke (combined, incident and prevalent)	CHARGE & SCG	NA	NA	NA	3.3×10^{-5}	NA	NA	NA	NA	3,017/19,540
Prevalent CAD	CARDIoGRAM	0.092	0.010	(b)	1.6×10^{-19}	28	1.29	1.38	(b)	22,233/64,726
Prevalent CAD	C4D ProCARDIS	0.132	0.022	(b)	2.2×10^{-9}	29	1.45	1.59	(b)	5,720/4,381
Prevalent CAD	C4D HPS	0.083	0.027	(b)	0.002	29	1.26	1.34	(b)	2,704/2,804
Prevalent CAD (combined)	CARDIoGRAM & C4D	0.100	0.009	(b)	8.1×10^{-29}	29	1.32	1.42	(b)	30,657/71,911
Prevalent chronic kidney disease	CKDGen	0.014	0.015	(b)	0.35	29	1.04	1.05	(b)	5,807/61,286
Prevalent microalbuminuria	CKDGen	0.008	0.019	(b)	0.68	29	1.02	1.03	(b)	3,698/27,882
Continuous measures of target organ damage										
Left ventricular mass (g)	EchoGen	0.822	0.317	(a)	0.01	29	2.30	2.89	(a)	12,612
Left ventricular wall thickness (cm)	EchoGen	0.009	0.002	(a)	6.0×10^{-6}	29	0.03	0.03	(a)	12,612
Serum creatinine	KidneyGen	-0.001	0.001	(d)	0.24	29	1.00	1.00	(d)	23,812
eGFR (four-parameter MDRD equation)	CKDGen	-0.0001	0.0009	(d)	0.93	29	1.00	1.00	(d)	67,093
Urinary albumin/creatinine ratio	CKDGen	0.005	0.007	(d)	0.43	29	1.01	1.02	(d)	31,580

Association of genetic risk score (using all 29 SNPs at 28 loci, parameterized using the average of SBP and DBP effects (= (SBP effect + DBP effect)/2) from the discovery analysis), tested in results from other GWAS consortia. (a) Units are the unit of phenotypic measurement, either per standard deviation (s.d.) of genetic risk score, or as a difference between top/bottom quintiles or deciles. (b) Units are ln(odds) per s.d. of genetic risk score, or odds ratio between top/bottom quintiles or deciles. (c) Units are ln(hazard) per s.d. of genetic risk score, or hazard ratio between top/bottom quintiles or deciles. (d) Units are ln(phenotype) per s.d. of genetic risk score, or phenotypic ratio between top/bottom quintiles or deciles. s.e., standard error. SCG, UK-US Stroke Collaborative Group; see Supplementary Materials sections 1.79 and 11 for further detail on consortia and studies.

derived from the 29 variants was significantly associated with blood-pressure-related organ damage and clinical cardiovascular disease, but not kidney disease. These loci improve our understanding of the genetic architecture of blood pressure, provide new biological insights into blood pressure control and may identify novel targets for the treatment of hypertension and the prevention of cardiovascular disease.

Note added in proof: Since this manuscript was submitted, Kato *et al.* published a blood pressure GWAS in East Asians that identified a SNP highly correlated to the SNP we report at the *NPR3/C5orf23* locus²⁸.

METHODS SUMMARY

Supplementary Materials provide complete methods and include the following sections: study recruitment and phenotyping, adjustment for antihypertensive medications, genotyping, data quality control, genotype imputation, within-cohort association analyses, meta-analyses of discovery and validation stages, stratified analyses by sex and BMI, identification of eSNPs and non-synonymous SNPs, metabolomic and lipidomic analyses, CNV analyses, pathway analyses, analyses for non-European ancestries, association of a risk score with hypertension and cardiovascular disease, estimation of numbers of undiscovered variants, measurement of natriuretic peptides, and brief literature reviews and GWAS database lookups of all validated blood pressure loci. Full GWAS results for ≈ 2.5 million SNPs are also provided.

Received 16 August 2010; accepted 28 July 2011.

Published online 11 September 2011.

- Levy, D. *et al.* Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the Framingham heart study. *Hypertension* **36**, 477–483 (2000).
- Kearney, P. M. *et al.* Global burden of hypertension: analysis of worldwide data. *Lancet* **365**, 217–223 (2005).
- Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* **360**, 1903–1913 (2002).
- Lifton, R. P., Gharavi, A. G. & Geller, D. S. Molecular mechanisms of human hypertension. *Cell* **104**, 545–556 (2001).
- Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genet.* **41**, 666–676 (2009).
- Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nature Genet.* **41**, 677–687 (2009).

- Meyer, T. E. *et al.* GOSR2 Lys67Arg is associated with hypertension in whites. *Am. J. Hypertens.* **22**, 163–168 (2009).
- Li, N. *et al.* Associations between genetic variations in the *FURIN* gene and hypertension. *BMC Med. Genet.* **11**, 124 (2010).
- Mussig, K. *et al.* 17 α -hydroxylase/17,20-lyase deficiency caused by a novel homozygous mutation (Y27Stop) in the cytochrome CYP17 gene. *J. Clin. Endocrinol. Metab.* **90**, 4362–4365 (2005).
- Ridker, P. M. *et al.* Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. *Clin. Chem.* **54**, 249–255 (2008).
- Burt, V. L. *et al.* Trends in the prevalence, awareness, treatment, and control of hypertension in the adult US population. Data from the health examination surveys, 1960 to 1991. *Hypertension* **26**, 60–69 (1995).
- Blood Pressure Lowering Treatment Trialists' Collaboration Effects of different regimens to lower blood pressure on major cardiovascular events in older and younger adults: meta-analysis of randomised trials. *Br. Med. J.* **336**, 1121–1123 (2008).
- Law, M. R., Morris, J. K. & Wald, N. J. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. *Br. Med. J.* **338**, b1665 (2009).
- Lewis, J. B. Blood pressure control in chronic kidney disease: is less really more? *J. Am. Soc. Nephrol.* **21**, 1086–1092 (2010).
- Park, J. H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature Genet.* **42**, 570–575 (2010).
- Newton-Cheh, C. *et al.* Association of common variants in *NPPA* and *NPPB* with circulating natriuretic peptides and blood pressure. *Nature Genet.* **41**, 348–353 (2009).
- Schenk, D. B. *et al.* Purification and subunit composition of atrial natriuretic peptide receptor. *Proc. Natl Acad. Sci. USA* **84**, 1521–1525 (1987).
- Schmidt, H. H. & Walter, U. NO at work. *Cell* **78**, 919–925 (1994).
- Matsukawa, N. *et al.* The natriuretic peptide clearance receptor locally modulates the physiological effects of the natriuretic peptide system. *Proc. Natl Acad. Sci. USA* **96**, 7403–7408 (1999).
- Friebe, A., Mergia, E., Dangel, O., Lange, A. & Koesling, D. Fatal gastrointestinal obstruction and hypertension in mice lacking nitric oxide-sensitive guanylyl cyclase. *Proc. Natl Acad. Sci. USA* **104**, 7699–7704 (2007).
- Ishimitsu, T., Ono, H., Minami, J. & Matsuoka, H. Pathophysiologic and therapeutic implications of adrenomedullin in cardiovascular disorders. *Pharmacol. Ther.* **111**, 909–927 (2006).
- Pushkin, A. *et al.* Cloning, tissue distribution, genomic organization, and functional characterization of NBC3, a new member of the sodium bicarbonate cotransporter family. *J. Biol. Chem.* **274**, 16569–16575 (1999).
- Hinkes, B. *et al.* Positional cloning uncovers mutations in *PLCE1* responsible for a nephrotic syndrome variant that may be reversible. *Nature Genet.* **38**, 1397–1405 (2006).

24. Feder, J. N. *et al.* A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nature Genet.* **13**, 399–408 (1996).
25. He, L., Wang, B., Hay, E. B. & Nebert, D. W. Discovery of ZIP transporters that participate in cadmium damage to testis and kidney. *Toxicol. Appl. Pharmacol.* **238**, 250–257 (2009).
26. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
27. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genet.* **42**, 937–948 (2010).
28. Kato, N. *et al.* Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nature Genet.* **43**, 531–538 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A number of the participating studies and authors are members of the CHARGE and Global BPgen consortia. Many funding mechanisms by NIH/NHLBI, European and private funding agencies contributed to this work and a full list is provided in section 21 of the Supplementary Materials.

Author Contributions Full author contributions and roles are listed in Supplementary Materials section 19.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.C. (aravinda@jhmi.edu), M.C. (m.j.caulfield@qmul.ac.uk), D.L. (levy@nhlbi.nih.gov), P.B.M. (p.b.munroe@qmul.ac.uk), C.N.-C. (cnewtonch@chgr.mgh.harvard.edu).

George B. Ehret^{1,2,3*}, Patricia B. Munroe^{4*}, Kenneth M. Rice^{5*}, Murielle Bochud^{2*}, Andrew D. Johnson^{6,7*}, Daniel I. Chasman^{8,9*}, Albert V. Smith^{10,11*}, Martin D. Tobin¹², Germaine C. Verwoert^{13,14,15}, Shih-Jen Hwang^{6,7,16}, Vasyli Pihur¹, Peter Vollenweider¹⁷, Paul F. O'Reilly¹⁸, Najaf Amin¹³, Jennifer L. Bragg-Gresham¹⁹, Alexander Teumer²⁰, Nicole L. Glazer²¹, Lenore Launer²², Jing Hua Zhao²³, Yuri Aulchenko¹³, Simon Heath²⁴, Sigmund Söber²⁵, Afshin Parsa²⁶, Jian'an Luan²³, Pankaj Arora²⁷, Abbas Dehghan^{13,14,15}, Feng Zhang²⁸, Gavin Lucas²⁹, Andrew A. Hicks³⁰, Anne U. Jackson³¹, John F. Peden³², Toshiko Tanaka³³, Sarah H. Wild³⁴, Igor Rudan^{35,36}, Wilmar Igl³⁷, Yuri Milaneschi³⁸, Alex N. Parker³⁸, Cristiano Fava^{39,40}, John C. Chambers^{18,41}, Ervin R. Fox⁴², Meena Kumari⁴³, Min Jin Go⁴⁴, Pim van der Harst⁴⁵, Wen Hong Linda Kao⁴⁶, Marketa Sjögren³⁹, D. G. Vinay⁴⁷, Myriam Alexander⁴⁸, Yasuharu Tabara⁴⁹, Sue Shaw-Hawkins⁵⁰, Peter H. Whincup⁵⁰, Yongmei Liu⁵¹, Gang Shi⁵², Johanna Kuusisto⁵³, Bamidele Tayo⁵⁴, Mark Seielstad^{55,56}, Xuelling Sim⁵⁷, Khanh-Dung Hoang Nguyen¹, Terho Lehtimäki⁵⁸, Giuseppe Matullo^{59,60}, Ying Wu⁶¹, Tom R. Gaunt⁶², N. Charlotte Omland-Orsted^{63,64}, Matthew N. Cooper⁶⁵, Carl G. P. Platou⁶⁶, Elin Org⁶⁵, Rebecca Hardy⁶⁷, Santosh Dahgam⁶⁸, Jutta Palmer⁶⁹, Veronique Vitart⁷⁰, Peter S. Braund^{71,72}, Tatiana Kuznetsova⁷³, Cuno S. P. M. Uiterwaal⁶³, Adebawale Adeyemo⁷⁴, Walter Palmas⁷⁵, Harry Campbell³⁵, Barbara Ludwig⁷⁶, Maciej Tomaszewski^{77,78}, Ioanna Tzoulaki^{77,78}, Nicholette D. Palmer⁷⁹, CARDIoGRAM consortium⁸⁰, CKDGen Consortium⁸¹, KidneyGen Consortium⁸², EchoGen consortium⁸³, CHARGE-HF consortium⁸⁴, Thor Aspelund^{10,11}, Melissa Garcia²², Yen-Pei C. Chang²⁶, Jeffrey R. O'Connell²⁶, Nanette I. Steinle²⁶, Diederick E. Grobbee⁸⁵, Dan E. Arking⁸⁶, Sharon L. Kardina⁸⁰, Alanna C. Morrison⁸¹, Dena Hernandez⁸², Samer Najjar^{83,84}, Wendy L. McArdle⁸⁵, David Hadley^{50,86}, Morris J. Brown⁸⁷, John M. Connell⁸⁸, Aaron D. Hingorani⁸⁹, Ian N. M. Day⁶², Debbie A. Lawlor⁶², John P. Beilby^{90,91}, Robert W. Lawrence⁶⁵, Robert Clarke⁹², Jemma C. Hopewell⁹², Halit Ongen³², Albert W. Dreisbach⁴², Yali Li⁹³, J. Hunter Young⁹⁴, Joshua C. Bis²¹, Mika Kahonen⁹⁵, Jorma Viikari⁹⁶, Linda S. Adair⁹⁷, Nanette R. Lee⁹⁸, Ming-Huei Chen⁹⁹, Matthias Olden^{100,101}, Cristian Pattaro³⁰, Judith A. Hoffman Bolton¹⁰², Anna Köttgen^{102,103}, Sven Bergmann^{104,105}, Vincent Mooser¹⁰⁶, Nish Chaturvedi¹⁰⁷, Timothy M. Frayling¹⁰⁸, Muhammad Islam¹⁰⁹, Tazeen H. Jafar¹⁰⁹, Jeanette Erdmann¹¹⁰, Smita R. Kulkarni¹¹¹, Stefan R. Bornstein⁷⁶, Jürgen Grässler⁷⁶, Leif Groop^{112,113}, Benjamin F. Voight¹¹⁴, Johannes Kettunen^{115,116}, Philip Howard¹¹⁷, Andrew Taylor⁴³, Simonetta Guarrera⁶⁰, Fulvio Ricceri^{59,60}, Valur Emilsson¹¹⁸, Andrew Plump¹¹⁸, Inês Barroso^{119,120}, Kay-Tee Khaw⁴⁸, Alan B. Weder¹²¹, Steven C. Hunt¹²², Yan V. Sun⁸⁰, Richard N. Bergman¹²³, Francis S. Collins¹²⁴, Lori L. Bonnycastle¹²⁴, Laura J. Scott³¹, Heather M. Stringham³¹, Leena Peltonen^{116,119,125,126}, Markus Perola¹²⁵, Erkki Vartiainen¹²⁵, Stefan-Martin Brand^{127,128}, Jan A. Staessen⁷³, Thomas J. Wang^{6,129}, Paul R. Burton^{127,128}, Maria Soler Artigas¹², Yanbin Dong¹³⁰, Harold Snieder^{130,131}, Xiaoling Wang¹³⁰, Haidong Zhu¹³⁰, Kurt K. Lohman¹³², Megan E. Rudock³¹, Susan R. Heckbert^{133,134}, Nicholas L. Smith^{133,134,135}, Kerri L. Wiggins¹³⁶, Ayo Doughty⁷⁴, Daniel Shrier⁷⁴, Gudrun Veldre^{25,137}, Margus Viigimaa^{138,139}, Sanjay Kinra¹⁴⁰, Dorairaj Prabhakaran¹⁴¹, Vikal Tripathy¹⁴¹, Carl D. Langefeld⁷⁹, Annika Rosengren¹⁴², Dag S. Thelle¹⁴³, Anna Maria Corsi¹⁴⁴, Andrew Singleton⁸², Terrence Forrester¹⁴⁵, Gina Hilton¹, Colin A. McKenzie¹⁴⁵, Tunde Salako¹⁴⁶, Naoharu Iwai¹⁴⁷, Yoshikuni Kita¹⁴⁸, Toshio Ogihara¹⁴⁹, Takayoshi Ohkubo^{148,150}, Tomonori Okamura^{147,148}, Hirotsugu Ueshima^{148,151}, Satoshi Umemura¹⁵², Susana Eyheramendy¹⁵³, Thomas Meitinger^{154,155}, H.-Erich Wichmann^{156,157,158}, Yoon Shin Cho⁴⁴, Hyung-Lae Kim⁴⁴, Jong-Young Lee⁴⁴, James Scott¹⁵⁹, Joban S. Sehmi^{41,159}, Weihua Zhang¹⁶⁰, Bo Hedblad³⁹, Peter Nilsson³⁹, George Davey Smith⁶², Andrew Wong⁶⁷, Narisu Narisu¹²⁴, Alena Stančáková⁵³, Leslie J. Raffel¹⁶⁰, Jie Yao¹⁶⁰, Sekar Kathiresan^{27,161}, Christopher J. O'Donnell^{9,27,162}, Stephen M. Schwartz¹³³, M. Arfan Ikram^{13,15}, W. T. Longstreth Jr¹⁶³, Thomas H. Mosley¹⁶⁴, Sudha Seshadri¹⁶⁵, Nick R.G. Shrine¹², Louise V. Wain¹²,

Mario A. Morken¹²⁴, Amy J. Swift¹²⁴, Jaana Laitinen¹⁶⁶, Inga Prokopenko^{51,167}, Paavo Zitting¹⁶⁸, Jackie A. Cooper⁶⁹, Steve E. Humphries⁶⁹, John Danesh⁴⁸, Asif Rasheed¹⁶⁹, Anuj Goel³², Anders Hamsten¹⁷⁰, Hugh Watkins³², Stephan J. L. Bakker¹⁷¹, Wiek H. van Gilst⁴⁵, Charles S. Janipalli⁴⁷, K. Radha Mani⁴⁷, Chittaranjan S. Yajnik¹¹¹, Albert Hofman¹³, Francesco U. S. Mattace-Raso^{13,14}, Ben A. Oostra¹⁷², Ayse Demirkan¹³, Aaron Isaacs¹³, Fernando Rivadeneira^{13,14}, Edward G. Lakatta¹⁷³, Marco Orru^{174,175}, Angelo Scuteri¹⁷³, Mika Ala-Korpela^{176,177,178}, Antti J. Kangas¹⁷⁶, Leo-Pekka Lytikäinen¹⁷⁸, Pasi Soininen^{176,177}, Taru Tuikainen^{176,179,180}, Peter Würtz^{181,176,179}, Rick Twee-Hee Ong^{56,57,181}, Marcus Dörner¹⁸², Heyo K. Kroemer¹⁸³, Uwe Völker²⁰, Henry Völzke¹⁸⁴, Pilar Galan¹⁸⁵, Serge Hercberg¹⁸⁵, Mark Lathrop²⁴, Diana Zelenika²⁴, Panos Deloukas¹¹⁹, Massimo Mangino²⁸, Tim D. Spector²⁸, Guangju Zhai²⁸, James F. Meschia¹⁸⁶, Michael A. Nalls⁸², Pankaj Sharma¹⁸⁷, Janos Terzic¹⁸⁸, M. V. Kranthi Kumar⁴⁷, Matthew Denniff⁷¹, Ewa Zukowska-Szczzechowska¹⁸⁹, Lynne E. Wagenknecht⁷⁹, F. Gerald R. Fowkes¹⁹⁰, Fadi J. Charchar¹⁹¹, Peter E. H. Schwarz¹⁹², Caroline Hayward⁷⁰, Xiuqing Guo¹⁶⁰, Charles Rotimi⁷⁴, Michiel L. Bots⁶³, Eva Brand¹⁹³, Nilesh J. Samani^{17,72}, Ozren Polasek¹⁹⁴, Philippa J. Talmud⁶⁰, Fredrik Nyberg^{68,195}, Diana Kuh⁶⁷, Maris Laan²⁵, Kristian Hveem⁶⁶, Lyle J. Palmer^{196,197}, Yvonne T. van der Schouw⁶³, Juan P. Casas¹⁹⁸, Karen L. Mohlke⁶¹, Paolo Vineis^{60,199}, Olli Raitakari²⁰⁰, Santhi K. Ganesh²⁰¹, Tien Y. Wong^{202,203}, E. Shyong Tai^{57,204,205}, Richard S. Cooper⁵⁴, Markku Laakso⁵³, Dabeeru C. Rao²⁰⁶, Tamara B. Harris²², Richard W. Morris²⁰⁷, Anna F. Dominiczak²⁰⁸, Mika Kivimäki²⁰⁹, Michael G. Marmot²⁰⁹, Tetsuro Miki⁴⁹, Danish Saleheen^{48,169}, Giriraj R. Chandak⁴⁷, Josef Coresh²¹⁰, Gerjan Navis²¹¹, Veikko Salomaa¹²⁵, Bok-Ghee Han⁴⁴, Xiaofeng Zhu⁹³, Jaspal S. Kooner^{41,159}, Olle Melander³⁹, Paul M. Ridker^{8,9,212}, Stefania Bandinelli²¹³, Ulf B. Gyllenstein³⁷, Alan F. Wright⁷⁰, James F. Wilson³⁴, Luigi Ferrucci³³, Martin Farrall³², Jaakko Tuomilehto^{214,215,216,217}, Peter P. Pramstaller^{30,218}, Roberto Elosua^{29,219}, Nicole Soranzo^{28,119}, Eric J. G. Sijbrands^{13,14}, David Altshuler^{114,220}, Ruth J. F. Loos²³, Alan R. Shuldiner^{26,221}, Christian Gieger¹⁵⁶, Pierre Meneton²²², Andre G. Uitterlinden^{13,14,15}, Nicholas J. Wareham²³, Vilmondur Gudnason^{10,11}, Jerome I. Rotter¹⁶⁰, Rainer Rettig²²³, Manuela Uda¹⁷⁴, David P. Strachan⁵⁰, Jacqueline C. M. Witterman^{13,15}, Anna-Liisa Hartikainen²²⁴, Jacques S. Beckmann^{104,225}, Eric Boerwinkle²²⁶, Ramachandran S. Vasan^{6,227}, Michael Boehnke³¹, Martin G. Larson^{6,228}, Marjo-Riitta Jarvelin^{18,229,230,231,232}, Bruce M. Psaty^{21,134*}, Gonçalo R. Abecasis^{19*}, Aravinda Chakravarti^{1*}, Paul Elliott^{18,232*}, Cornelia M. van Duijn^{13,233*}, Christopher Newton-Cheh^{27,114*}, Daniel Levy^{6,7,16*}, Mark J. Caulfield^{4*} & Toby Johnson^{4*}

¹Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ²Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois and University of Lausanne, Bugnon 17, 1005 Lausanne, Switzerland. ³Cardiology, Department of Specialties of Internal Medicine, Geneva University Hospital, Rue Gabrielle-Perret-Gentil 4, 1211 Geneva 14, Switzerland. ⁴Clinical Pharmacology and The Genome Centre, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. ⁵Department of Biostatistics, University of Washington, Seattle, Washington 98195, USA. ⁶Framingham Heart Study, Framingham, Massachusetts 01702, USA. ⁷National Heart Lung, and Blood Institute, Bethesda, Maryland 20824, USA. ⁸Division of Preventive Medicine, Brigham and Women's Hospital, 900 Commonwealth Avenue East, Boston, Massachusetts 02215, USA. ⁹Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁰Icelandic Heart Association, 201 Kópavogur, Iceland. ¹¹University of Iceland, 101 Reykjavik, Iceland. ¹²Department of Health Sciences, University of Leicester, University Rd, Leicester LE1 7RH, UK. ¹³Department of Epidemiology, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands. ¹⁴Department of Internal Medicine, Erasmus Medical Center, 3000 CA Rotterdam, The Netherlands. ¹⁵Netherlands Consortium for Healthy Aging (NCHA), Netherlands Genome Initiative (NGI), Erasmus 3000 CA Rotterdam, The Netherlands. ¹⁶Center for Population Studies, National Heart Lung, and Blood Institute, Bethesda, Maryland 20824, USA. ¹⁷Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland. ¹⁸Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK. ¹⁹Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan 48103, USA. ²⁰Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany. ²¹Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, Washington 98101, USA. ²²Laboratory of Epidemiology, Demography, Biometry, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. ²³MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge CB2 0QQ, UK. ²⁴Centre National de Génétique, Commissariat à l'Energie Atomique, Institut de Génétique, 91057 Evry, France. ²⁵Institute of Molecular and Cell Biology, University of Tartu, Riia 23, Tartu 51010, Estonia. ²⁶University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. ²⁷Center for Human Genetic Research, Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ²⁸Department of Twin Research & Genetic Epidemiology, King's College London, London SE1 7EH, UK. ²⁹Cardiovascular Epidemiology and Genetics, Institut Municipal d'Investigació Mèdica, Barcelona Biomedical Research Park, 88 Doctor Aiguader, 08003 Barcelona, Spain. ³⁰Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Viale Druso 1, 39100 Bolzano, Italy - Affiliated Institute of the University of Lübeck, Germany. ³¹Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA. ³²Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. ³³Clinical Research Branch, National Institute on Aging, Baltimore, Maryland 21205, USA. ³⁴Centre for Population Health Sciences, University of Edinburgh, EH8 9AG, UK. ³⁵Centre for Population Health Sciences and Institute of Genetics and Molecular Medicine, College of Medicine and Veterinary Medicine, University of Edinburgh, EH8 9AG, UK. ³⁶Croatian Centre for Global Health,

University of Split, 21000 Split, Croatia.³⁷Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, SE-751 85 Uppsala, Sweden.³⁸Amgen, 1 Kendall Square, Building 100, Cambridge, Massachusetts 02139, USA.³⁹Department of Clinical Sciences, Lund University, 205 02 Malmö, Sweden.⁴⁰Department of Medicine, University of Verona, 37134 Verona, Italy.⁴¹Ealing Hospital, London UB1 3HJ, UK.⁴²Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA.⁴³Genetic Epidemiology Group, Epidemiology and Public Health, UCL, London, WC1E 6BT, UK.⁴⁴Center for Genome Science, National Institute of Health, Seoul 122-701, Korea.⁴⁵Department of Cardiology, University Medical Center Groningen, University of Groningen, 9713 GZ Groningen, The Netherlands.⁴⁶Departments of Epidemiology and Medicine, Johns Hopkins University, Baltimore, Maryland 21205, USA.⁴⁷Centre for Cellular and Molecular Biology (CCMB), Council of Scientific and Industrial Research (CSIR), Uppal Road, Hyderabad 500 007, India.⁴⁸Department of Public Health and Primary Care, University of Cambridge, CB1 8RN, UK.⁴⁹Department of Basic Medical Research and Education, and Department of Geriatric Medicine, Ehime University Graduate School of Medicine, Toon, 791-0295, Japan.⁵⁰Division of Community Health Sciences, St George's University of London, London SW17 0RE, UK.⁵¹Epidemiology & Prevention, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, USA.⁵²Division of Biostatistics and Department of Genetics, School of Medicine, Washington University in St. Louis, Saint Louis, Missouri 63110, USA.⁵³Department of Medicine, University of Eastern Finland and Kuopio University Hospital, 70210 Kuopio, Finland.⁵⁴Department of Preventive Medicine and Epidemiology, Loyola University Medical School, Maywood, Illinois 60153, USA.⁵⁵Department of Laboratory Medicine & Institute of Human Genetics, University of California San Francisco, 513 Parnassus Ave. San Francisco, California 94143, USA.⁵⁶Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672, Singapore.⁵⁷Centre for Molecular Epidemiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore.⁵⁸Department of Clinical Chemistry, University of Tampere and Tampere University Hospital, Tampere 33521, Finland.⁵⁹Department of Genetics, Biology and Biochemistry, University of Torino, Via Santena 19, 10126 Torino, Italy.⁶⁰Human Genetics Foundation (HUGF), Via Nizza 52, 10126 Torino, Italy.⁶¹Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA.⁶²MRC Centre for Causal Analyses in Translational Epidemiology, School of Social & Community Medicine, University of Bristol, Bristol BS8 2BN, UK.⁶³Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Heidelberglaan 100, 3508 GA Utrecht, The Netherlands.⁶⁴Complex Genetics Section, Department of Medical Genetics - DBG, University Medical Center Utrecht, 3508 GA Utrecht, The Netherlands.⁶⁵Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, Crawley, Western Australia 6009, Australia.⁶⁶HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, 7600 Levanger, Norway.⁶⁷MRC Unit for Lifelong Health & Ageing, London WC1B 5JU, UK.⁶⁸Occupational and Environmental Medicine, Department of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, 40530 Gothenburg, Sweden.⁶⁹Centre for Cardiovascular Genetics, University College London, London WC1E 6JF, UK.⁷⁰MRC Human Genetics Unit and Institute of Genetics and Molecular Medicine, Edinburgh EH2, UK.⁷¹Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester LE3 9QP, UK.⁷²Leicester NIHR Biomedical Research Unit in Cardiovascular Disease, Glenfield Hospital, Leicester, LE3 9QP, UK.⁷³Studies Coordinating Centre, Division of Hypertension and Cardiac Rehabilitation, Department of Cardiovascular Diseases, University of Leuven, Campus Sint Rafaël, Kapucijnenvoer 35, Block D, Box 7001, 3000 Leuven, Belgium.⁷⁴Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, Maryland 20892, USA.⁷⁵Columbia University, New York, New York 10027, USA.⁷⁶Department of Medicine III, Medical Faculty Carl Gustav Carus at the Technical University of Dresden, 01307 Dresden, Germany.⁷⁷Epidemiology and Biostatistics, School of Public Health, Imperial College, London W2 1PG, UK.⁷⁸Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, 45110 Ioannina, Greece.⁷⁹Wake Forest University Health Sciences, Winston-Salem, North Carolina 27157, USA.⁸⁰Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan 48109, USA.⁸¹Division of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, University of Texas at Houston Health Science Center, 12 Herman Pressler, Suite 453E, Houston, Texas 77030, USA.⁸²Laboratory of Neurogenetics, National Institute on Aging, Bethesda, Maryland 20892, USA.⁸³Laboratory of Cardiovascular Science, Intramural Research Program, National Institute on Aging, NIH, Baltimore, Maryland 21224, USA.⁸⁴Washington Hospital Center, Division of Cardiology, Washington, District of Columbia 20010, USA.⁸⁵ALSPAC Laboratory, University of Bristol, Bristol BS8 2BN, UK.⁸⁶Pediatric Epidemiology Center, University of South Florida, Tampa, Florida 33612, USA.⁸⁷Clinical Pharmacology Unit, University of Cambridge, Addenbrookes Hospital, Hills Road, Cambridge CB2 2QQ, UK.⁸⁸University of Dundee, Ninewells Hospital & Medical School, Dundee DD1 9SY, UK.⁸⁹Genetic Epidemiology Group, Department of Epidemiology and Public Health, UCL, London WC1E 6BT, UK.⁹⁰Pathology and Laboratory Medicine, University of Western Australia, Crawley, Western Australia 6009, Australia.⁹¹Molecular Genetics, PathWest Laboratory Medicine, Nedlands, Western Australia 6009, Australia.⁹²Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford OX3 7LF, UK.⁹³Department of Epidemiology and Biostatistics, Case Western Reserve University, 2103 Cornell Road, Cleveland, Ohio 44106, USA.⁹⁴Department of Medicine, Johns Hopkins University, Baltimore 21205, USA.⁹⁵Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, 33521, Finland.⁹⁶Department of Medicine, University of Turku and Turku University Hospital, Turku 20521, Finland.⁹⁷Department of Nutrition, University of North Carolina, Chapel Hill, North Carolina 27599, USA.⁹⁸Office of Population Studies Foundation, University of San Carlos, Talamban, Cebu City 6000, Philippines.⁹⁹Department of Neurology and Framingham Heart Study, Boston University School of Medicine, Boston, Massachusetts 02118, USA.¹⁰⁰Department of Internal Medicine II, University Medical Center Regensburg, 93053 Regensburg, Germany.¹⁰¹Department of Epidemiology and Preventive Medicine, University Medical Center Regensburg, 93053 Regensburg, Germany.¹⁰²Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland 21205, USA.¹⁰³Renal Division, University Hospital Freiburg, 79095 Freiburg, Germany.¹⁰⁴Département de Génétique Médicale, Université de Lausanne, 1015 Lausanne, Switzerland.¹⁰⁵Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland.¹⁰⁶Division of Genetics, GlaxoSmithKline, Philadelphia, Pennsylvania 19101, USA.¹⁰⁷International Centre for Circulatory Health, National Heart & Lung Institute, Imperial College, London SW7 2AZ, UK.¹⁰⁸Genetics of Complex Traits, Peninsula Medical School, University of Exeter, Exeter EX4 4QJ, UK.¹⁰⁹Department of Community Health Sciences & Department of Medicine, Aga Khan University, Karachi 74800, Pakistan.¹¹⁰Medizinische Klinik II, Universität zu Lübeck, 23538 Lübeck, Germany.¹¹¹Diabetes Unit, KEM Hospital and Research Centre, Rasta Peth, Pune-411011, Maharashtra, India.¹¹²Department of Clinical Sciences, Diabetes and Endocrinology Research Unit, University Hospital, 205 02 Malmö, Sweden.¹¹³Lund University, Malmö 20502, Sweden.¹¹⁴Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02139, USA.¹¹⁵Department of Chronic Disease Prevention, National Institute for Health and Welfare, 00251 Helsinki, Finland.¹¹⁶FIMM, Institute for Molecular Medicine, Finland, Biomedicum, P.O. Box 104, 00251 Helsinki, Finland.¹¹⁷William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK.¹¹⁸Merck Research Laboratory, 126 East Lincoln Avenue, Rahway, New Jersey 07065, USA.¹¹⁹Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK.¹²⁰University of Cambridge Metabolic Research Labs, Institute of Metabolic Science Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.¹²¹Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA.¹²²Cardiovascular Genetics, University of Utah School of Medicine, Salt Lake City, Utah 84132, USA.¹²³Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA.¹²⁴National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.¹²⁵National Institute for Health and Welfare, 00271 Helsinki, Finland.¹²⁶Broad Institute, Cambridge, Massachusetts 02142, USA.¹²⁷Leibniz-Institute for Arteriosclerosis Research, Department of Molecular Genetics of Cardiovascular Disease, University of Münster, 48149 Münster, Germany.¹²⁸Medical Faculty of the Westfälische Wilhelms University Muenster, Department of Molecular Genetics of Cardiovascular Disease, University of Münster, 48149 Münster, Germany.¹²⁹Division of Cardiology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.¹³⁰Georgia Prevention Institute, Department of Pediatrics, Medical College of Georgia, Augusta, Georgia 30912, USA.¹³¹Unit of Genetic Epidemiology and Bioinformatics, Department of Epidemiology, University Medical Center Groningen, University of Groningen, 9713 GZ Groningen, The Netherlands.¹³²Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, USA.¹³³Department of Epidemiology, University of Washington, Seattle, Washington 98195, USA.¹³⁴Group Health Research Institute, Group Health Cooperative, Seattle, Washington 98124, USA.¹³⁵Seattle Epidemiologic Research and Information Center, Veterans Health Administration Office of Research & Development, Seattle, Washington 98108, USA.¹³⁶Department of Medicine, University of Washington, Seattle, Washington 98195, USA.¹³⁷Department of Cardiology, University of Tartu, L. Puusepa 8, 51014 Tartu, Estonia.¹³⁸Tallinn University of Technology, Institute of Biomedical Engineering, Ehitajate tee 5, 19086 Tallinn, Estonia.¹³⁹Centre of Cardiology, North Estonia Medical Centre, Sõistete tee 19, 13419 Tallinn, Estonia.¹⁴⁰Department of Non-communicable disease Epidemiology, The London School of Hygiene and Tropical Medicine London, Keppel Street, London WC1E 7HT, UK.¹⁴¹South Asia Network for Chronic Disease, Public Health Foundation of India, C-1/52, SDA, New Delhi 100016, India.¹⁴²Department of Emergency and Cardiovascular Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, 41685 Gothenburg, Sweden.¹⁴³Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, 0317 Oslo, Norway.¹⁴⁴Tuscany Regional Health Agency, 50129 Florence, Italy.¹⁴⁵Tropical Medicine Research Institute, University of the West Indies, Mona, Kingston, Jamaica.¹⁴⁶University of Ibadan, 200284 Ibadan, Nigeria.¹⁴⁷Department of Genomic Medicine, and Department of Preventive Cardiology, National Cerebral and Cardiovascular Research Center, Suita, 565-8565, Japan.¹⁴⁸Department of Health Science, Shiga University of Medical Science, Otsu, 520-2192, Japan.¹⁴⁹Department of Geriatric Medicine, Osaka University Graduate School of Medicine, Suita, 565-0871, Japan.¹⁵⁰Tohoku University Graduate School of Pharmaceutical Sciences and Medicine, Sendai, 980-8578, Japan.¹⁵¹Lifestyle-related Disease Prevention Center, Shiga University of Medical Science, Otsu, 520-2192, Japan.¹⁵²Department of Medical Science and Cardiorenal Medicine, Yokohama City University School of Medicine, Yokohama, 236-0004, Japan.¹⁵³Department of Statistics, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile.¹⁵⁴Institute of Human Genetics, Helmholtz Zentrum Munich, German Research Centre for Environmental Health, 85764 Neuherberg, Germany.¹⁵⁵Institute of Human Genetics, Klinikum rechts der Isar, Technical University of Munich, 81675 Munich, Germany.¹⁵⁶Institute of Epidemiology, Helmholtz Zentrum Munich, German Research Centre for Environmental Health, 85764 Neuherberg, Germany.¹⁵⁷Chair of Epidemiology, Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, 81377 Munich, Germany.¹⁵⁸Klinikum Grosshadern, 81377 Munich, Germany.¹⁵⁹National Heart and Lung Institute, Imperial College London, London W12 0HS, UK.¹⁶⁰Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California 90048, USA.¹⁶¹Medical Population Genetics, Broad Institute of Harvard and MIT, 5 Cambridge Center, Cambridge, Massachusetts 02142, USA.¹⁶²National Heart, Lung and Blood Institute and its Framingham Heart Study, 73 Mount Wayte Ave., Suite #2, Framingham, Massachusetts 01702, USA.¹⁶³Department of Neurology and Medicine, University of Washington, Seattle, Washington 98195, USA.¹⁶⁴Department of Medicine (Geriatrics), University of Mississippi Medical Center, Jackson, Mississippi 39216, USA.¹⁶⁵Department of Neurology, Boston University School of Medicine, Massachusetts 02118, USA.¹⁶⁶Finnish Institute of Occupational Health, Aapistie 1, 90220 Oulu, Finland.¹⁶⁷Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.¹⁶⁸Lapland Central Hospital, Department of Physiatrics, Box 8041, 96101 Rovaniemi, Finland.¹⁶⁹Center for

- Non-Communicable Diseases Karachi 74800, Pakistan.¹⁷⁰Atherosclerosis Research Unit, Department of Medicine, Karolinska Institute, 171 77 Stockholm, Sweden.
- ¹⁷¹Department of Internal Medicine, University Medical Center Groningen, University of Groningen, 9713 GZ Groningen, The Netherlands.¹⁷²Department of Clinical Genetics, Erasmus Medical Center, 3000 CA Rotterdam, The Netherlands.¹⁷³Gerontology Research Center, National Institute on Aging, Baltimore, Maryland 21224, USA.¹⁷⁴Istituto di Neurogenetica e Neurofarmacologia, Consiglio Nazionale delle Ricerche, Cittadella Universitaria di Monserrato, 09042 Monserrato, Cagliari, Italy.¹⁷⁵Unita Operativa Semplice Cardiologia, Divisione di Medicina, Presidio Ospedaliero Santa Barbara, 09016 Iglesias, Italy.¹⁷⁶Computational Medicine Research Group, Institute of Clinical Medicine, University of Oulu and Biocenter Oulu, 90014 University of Oulu, Oulu, Finland.¹⁷⁷NMR Metabonomics Laboratory, Department of Biosciences, University of Eastern Finland, 70211 Kuopio, Finland.¹⁷⁸Department of Internal Medicine and Biocenter Oulu, Clinical Research Center, 90014 University of Oulu, Oulu, Finland.¹⁷⁹Institute for Molecular Medicine Finland FIMM, 00014 University of Helsinki, Helsinki, Finland.¹⁸⁰Department of Biomedical Engineering and Computational Science, School of Science and Technology, Aalto University, 00076 Aalto, Espoo, Finland.¹⁸¹NUS Graduate School for Integrative Sciences & Engineering (NGS) Centre for Life Sciences (CeLS), Singapore 117456, Singapore.¹⁸²Department of Internal Medicine B, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany.¹⁸³Institute of Pharmacology, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany.¹⁸⁴Institute for Community Medicine, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany.¹⁸⁵U557 Institut National de la Santé et de la Recherche Médicale, U1125 Institut National de la Recherche Agronomique, Université Paris 13, 93017 Bobigny, France.¹⁸⁶Department of Neurology, Mayo Clinic, Jacksonville, Florida 32224, USA.¹⁸⁷Imperial College Cerebrovascular Unit (ICCRU), Imperial College, London W6 8RF, UK.¹⁸⁸Faculty of Medicine, University of Split, 21000 Split, Croatia.¹⁸⁹Department of Internal Medicine, Diabetology, and Nephrology, Medical University of Silesia, 41-800, Zabrze, Poland.¹⁹⁰Public Health Sciences section, Division of Community Health Sciences, University of Edinburgh, Medical School, Teviot Place, Edinburgh, EH8 9AG, UK.¹⁹¹School of Science and Engineering, University of Ballarat, 3353 Ballarat, Australia.¹⁹²Prevention and Care of Diabetes, Department of Medicine III, Medical Faculty Carl Gustav Carus at the Technical University of Dresden, 01307 Dresden, Germany.¹⁹³University Hospital Münster, Internal Medicine D, 48149 Münster, Germany.¹⁹⁴Department of Medical Statistics, Epidemiology and Medical Informatics, Andrija Stampar School of Public Health, University of Zagreb, 10000 Zagreb, Croatia.¹⁹⁵AstraZeneca R&D, 431 83 Mölndal, Sweden.¹⁹⁶Genetic Epidemiology & Biostatistics Platform, Ontario Institute for Cancer Research, Toronto, Ontario M5G 1L7, Canada.¹⁹⁷Samuel Lunenfeld Institute for Medical Research, University of Toronto, Toronto, Ontario M5S 1A1, Canada.¹⁹⁸Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.¹⁹⁹Department of Epidemiology and Public Health, Imperial College, Norfolk Place, London W2 1PG, UK.²⁰⁰Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and the Department of Clinical Physiology, Turku University Hospital, Turku, 20521, Finland.²⁰¹Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan Medical Center, Ann Arbor, Michigan 48109, USA.²⁰²Singapore Eye Research Institute, Singapore 168751, Singapore.²⁰³Department of Ophthalmology, National University of Singapore, Singapore 119074, Singapore.²⁰⁴Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119074, Singapore.
- ²⁰⁵Duke-National University of Singapore Graduate Medical School, Singapore 169857, Singapore.²⁰⁶Division of Biostatistics, Washington University School of Medicine, Saint Louis, Missouri 63110, USA.²⁰⁷Department of Primary Care & Population Health, UCL, London NW3 2PF, UK.²⁰⁸BHF Glasgow Cardiovascular Research Centre, University of Glasgow, 126 University Place, Glasgow G12 8TA, UK.²⁰⁹Epidemiology Public Health, UCL, London WC1E 6BT, UK.²¹⁰Departments of Epidemiology, Biostatistics, and Medicine, Johns Hopkins University, Baltimore, Maryland 21205, USA.²¹¹Division of Nephrology, Department of Internal Medicine, University Medical Center Groningen, University of Groningen, 9713 GZ Groningen, The Netherlands.²¹²Division of Cardiology, Brigham and Women's Hospital, 900 Commonwealth Avenue East, Boston, Massachusetts 02215, USA.²¹³Geriatric Rehabilitation Unit, Azienda Sanitaria Firenze (ASF), 50100 Florence, Italy.²¹⁴National Institute for Health and Welfare, Diabetes Prevention Unit, 00271 Helsinki, Finland.²¹⁵Hjelt Institute, Department of Public Health, University of Helsinki, 00014 Helsinki, Finland.²¹⁶South Ostrobothnia Central Hospital, 60220 Seinäjoki, Finland.²¹⁷Red RECAVA Grupo RD06/0014/0015, Hospital Universitario La Paz, 28046 Madrid, Spain.²¹⁸Department of Neurology, General Central Hospital, 39100 Bolzano, Italy.²¹⁹CIBER Epidemiología y Salud Pública, 08003 Barcelona, Spain.²²⁰Department of Medicine and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.²²¹Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, Maryland 21201, USA.²²²U872 Institut National de la Santé et de la Recherche Médicale, Centre de Recherche des Cordeliers, 75006 Paris, France.²²³Institute of Physiology, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany.²²⁴Institute of Clinical Medicine/Obstetrics and Gynecology, University of Oulu, 90014 Oulu, Finland.
- ²²⁵Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland.²²⁶Human Genetics Center, 1200 Hermann Pressler, Suite E447 Houston, Texas 77030, USA.²²⁷Division of Epidemiology and Prevention, Boston University School of Medicine, Boston, Massachusetts 02215, USA.²²⁸Department of Mathematics, Boston University, Boston, Massachusetts 02215, USA.²²⁹Institute of Health Sciences, University of Oulu, BOX 5000, 90014 University of Oulu, Finland.²³⁰Biocenter Oulu, University of Oulu, BOX 5000, 90014 University of Oulu, Finland.²³¹National Institute for Health and Welfare, Box 310, 90101 Oulu, Finland.²³²MRC-HPA Centre for Environment and Health, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK.²³³Centre of Medical Systems Biology (CMSB 1-2), NGL Erasmus Medical Center, Rotterdam, The Netherlands.

*These authors contributed equally to this work.

†A full list of authors and affiliations appears in Supplementary Information.

‡Deceased.

Primary forests are irreplaceable for sustaining tropical biodiversity

Luke Gibson^{1*}, Tien Ming Lee^{2,3*}, Lian Pin Koh^{1,4}, Barry W. Brook⁵, Toby A. Gardner⁶, Jos Barlow⁷, Carlos A. Peres⁸, Corey J. A. Bradshaw^{5,9}, William F. Laurance¹⁰, Thomas E. Lovejoy^{11,12} & Navjot S. Sodhi^{1‡}

Human-driven land-use changes increasingly threaten biodiversity, particularly in tropical forests where both species diversity and human pressures on natural environments are high¹. The rapid conversion of tropical forests for agriculture, timber production and other uses has generated vast, human-dominated landscapes with potentially dire consequences for tropical biodiversity^{2–5}. Today, few truly undisturbed tropical forests exist, whereas those degraded by repeated logging and fires, as well as secondary and plantation forests, are rapidly expanding^{6,7}. Here we provide a global assessment of the impact of disturbance and land conversion on biodiversity in tropical forests using a meta-analysis of 138 studies. We analysed 2,220 pairwise comparisons of biodiversity values in primary forests (with little or no human disturbance) and disturbed forests. We found that biodiversity values were substantially lower in degraded forests, but that this varied considerably by geographic region, taxonomic group, ecological metric and disturbance type. Even after partly accounting for confounding colonization and succession effects due to the composition of surrounding habitats, isolation and time since disturbance, we find that most forms of forest degradation have an overwhelmingly detrimental effect on tropical biodiversity. Our results clearly indicate that when it comes to maintaining tropical biodiversity, there is no substitute for primary forests.

As the extent of primary forests is shrinking throughout the tropics, a growing body of work has quantified the biodiversity values of degraded tropical forests. The ecological responses following forest conversion vary markedly across taxonomic groups, human impact types, ecological metrics and geographic regions^{5,8–10}. Most studies, however, provide limited insight into the varied responses of tropical forest biota to human impacts because they are understandably restricted to particular disturbance types^{11,12}, taxa^{13,14} and geographic regions¹⁵. Therefore, their often contrasting conclusions might have clouded ongoing debates over the conservation value of modified forest ecosystems⁴. A comprehensive meta-analysis of the conservation value of human-modified tropical forests is therefore sorely lacking. Notably, such an assessment could provide a critical baseline for monitoring progress towards global conservation targets¹⁶, evaluate the biodiversity benefits of international carbon-trading initiatives to reduce emissions from deforestation and forest degradation^{17,18} (for example the United Nations REDD+ programme), and guide policy development through the integration of biodiversity data into the modelling of land-use change scenarios^{2,19,20}.

Here we conduct a global meta-analysis to measure the varied effects of land-use change and forest degradation on biodiversity in tropical

forests. From an exhaustive literature search, we identified 138 studies that reported measures of biodiversity from multiple sites in both primary and disturbed tropical forests (Methods). We necessarily assumed that all 'primary forests' referred to in our source literature are largely old-growth forests that have experienced little to no recent human disturbance, although we recognize that in reality few primary forests are likely to be genuinely pristine. Primary forests are starkly differentiated from disturbed sites, which encompass the full spectrum of degraded and converted forest types, including selectively logged forests, secondary forests and forests converted into various forms of agriculture. In total, these studies spanned 28 countries and 92 study landscapes (Fig. 1). To measure the effect size of human-driven land-use changes, we calculated the weighted average of the standardized difference (based on pooled variance measures) between mean biodiversity measurements in primary and disturbed sites²¹ (that is, Hedges' g^*). The effect size was positive when the biodiversity value of primary forest sites was greater than that of disturbed sites, implying that the measured disturbance had a detrimental impact on biodiversity. We used a resampling procedure based on 10,000 bootstrap samples (with replacement) to generate the median effect size and 95% confidence intervals.

Overall, human impacts reduced biodiversity in tropical forests, although the effect size varied by region, taxonomic group, metric and disturbance type (Fig. 2). The median effect size for all 2,220 pairwise comparisons from 138 studies was 0.51 (95% confidence interval, 0.44–0.58) (Supplementary Table 1). This changed little when we accounted for pseudoreplication from studies that reported multiple comparisons, using a resampling procedure in which one comparison per study was randomly drawn for 10,000 samples, yielding an overall effect size of 0.57 (0.35–0.79) (Supplementary Table 1). Our results are also robust to publication biases (Methods). The surrounding habitat might either ameliorate (if hospitable) or exacerbate (if hostile) the impact of forest disturbance on biodiversity²². Although data are lacking for a comprehensive analysis, to account partly for this effect we repeated our analysis using only those studies that had natural vegetation (that is, primary and selectively logged forests) as the surrounding habitat (70.1% of all pairwise comparisons). Using this subset, we detected no substantial change in either the direction or the magnitude of effect sizes for the full data set (0.58, 0.49–0.68), or for each of the variables described below (Supplementary Table 1).

We found that human impacts on biodiversity varied by region. Although our data set is highly comprehensive, it is still limited given the vast extent of tropical forests and the myriad ways in which

¹Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore. ²Ecology, Behavior and Evolution Section, Division of Biological Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0116, USA. ³Department of Ecology and Evolutionary Biology, Yale University, PO Box 208106, New Haven, Connecticut 06520-8106, USA. ⁴Department of Environmental Sciences, ETH Zurich, CHN G73.1, Universitätsstrasse, 16 8092 Zurich, Switzerland. ⁵The Environment Institute and School of Earth and Environmental Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia. ⁶Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. ⁷Lancaster Environmental Centre, Lancaster University, Lancaster LA1 4YQ, UK. ⁸School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. ⁹South Australian Research and Development Institute, PO Box 120, Henley Beach, South Australia 5022, Australia. ¹⁰Centre for Tropical Environmental and Sustainability Science (TESS) and School of Marine and Tropical Biology, James Cook University, Cairns, Queensland 4870, Australia.

¹¹Biological Dynamics of Forest Fragments Project, National Institute for Amazonian Research (INPA), CP 478, Manaus, AM 69011-970, Brazil. ¹²H. John Heinz III Center for Science, Economics and Environment, 1001 Pennsylvania Avenue NW, Washington DC 20004, USA.

*These authors contributed equally to this work.

‡Deceased.

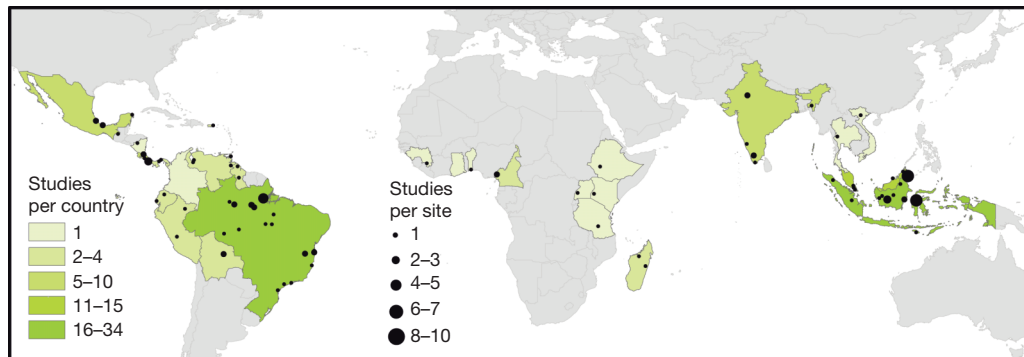


Figure 1 | Map of study sites by country and by study location. Country colour represents the number of studies per country ($n = 28$ total countries) and circle size represents the number of studies at each site ($n = 92$ total sites; only 82 sites with Global Positioning System coordinates are shown).

humans disturb them²³. Asia (52 studies) and South America (47) were the subjects of considerably more studies than were Central America (27) and Africa (12) (Fig. 1 and Supplementary Table 1). This regional bias implies that our findings might be more generalizable to Asia and South America than to other tropical regions. More critically, it highlights an urgent need for more research, particularly in Africa, which sustains the second largest contiguous tropical forest in the world⁵. Despite this important caveat, we found that Asia harbours the most sensitive biota, producing an effect size of 0.95 (0.83–1.08), which is substantially higher than that of the other three regions (Fig. 2a). This highlights the great toll human land-use changes are exacting in Asia, particularly in Southeast Asia, which most Asian studies (44 of 52) considered. Recent and widespread expansion of oil palm monoculture and exotic-tree plantations has greatly modified forest habitats in this region²⁴, but all forms of human impact were higher in Asia than

elsewhere (Fig. 3a), suggesting that this regional pattern holds regardless of disturbance type. Our results highlight the critical need to mitigate the particularly detrimental human impacts in Asia²⁵.

Most taxonomic groups we assessed were negatively affected by disturbance, with effect sizes greater than 0.5 (Fig. 2b and Supplementary Fig. 1b). However, mammals were less sensitive to the disturbances measured and, in some instances, actually benefitted from human disturbance, with an effect size of -0.12 (-0.24 to -0.01). This disparity, largely due to higher mammal abundances in certain disturbance types (Fig. 3b and Supplementary Table 3), might arise because of mammals' high tolerance of degraded forests and forest edges²⁶, particularly among small mammals (-0.04 , -0.27 to 0.20) and bats (-0.24 , -0.42 to -0.06), which dominated most studies on mammals (Supplementary Table 1). At the other extreme, birds were the most sensitive group, with an effect size of 0.72 (0.52–0.93).

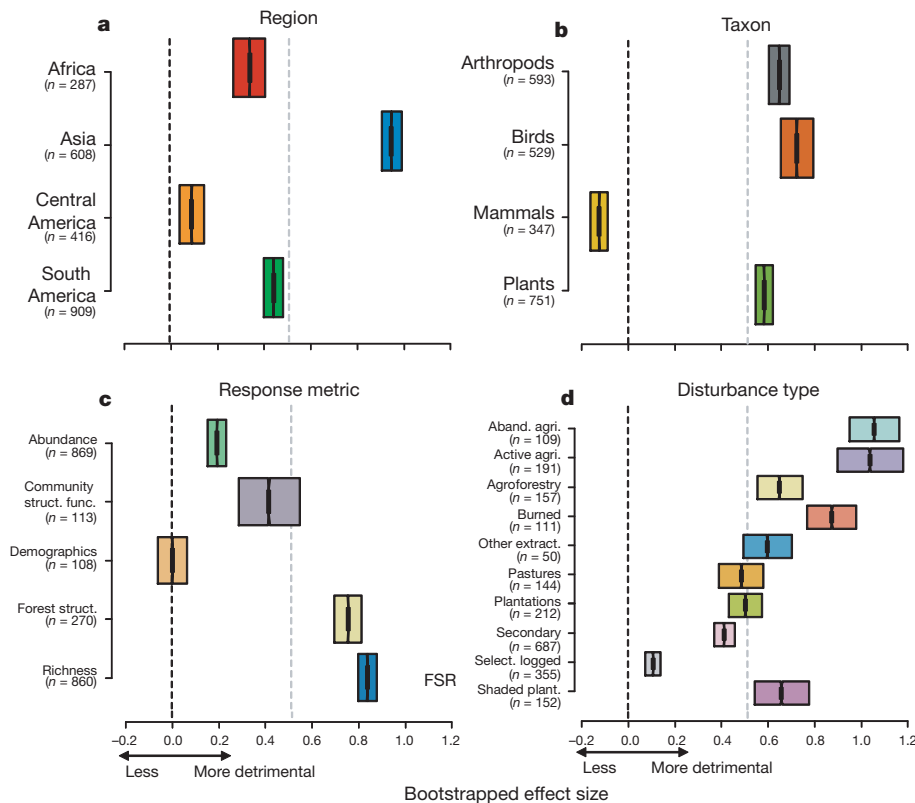


Figure 2 | Box plots of bootstrapped effect size. a, By region; b, by taxon; c, by response metric; d, by disturbance type (omitting clear-cut and disturbed/hunted owing to small sample sizes, that is, <50 comparisons). Plotted are median values and interquartile ranges of 10,000 resampled (with replacement) effect size calculations for each group. Widths of notches in box plots

approximate 95% confidence intervals. Median value for forest species richness (FSR) is plotted for comparison. The vertical black and grey dashed lines represent an effect size of zero and the median effect size for the entire data set, respectively. Sample size is shown in parentheses.

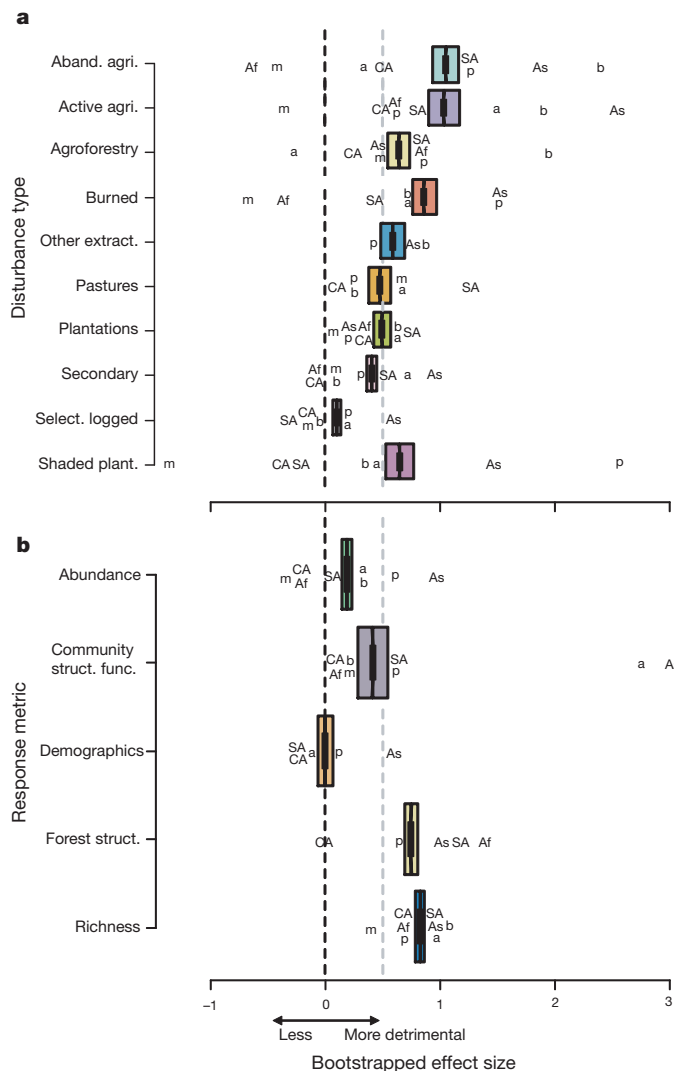


Figure 3 | Box plots of bootstrapped effect size. **a**, By disturbance type; **b**, by response metric, as in Fig. 2. Median effect size is also plotted as a function of region and taxon, with overlapping points stacked: Af, Africa; As, Asia; CA, Central America; SA, South America; a, arthropods; b, birds; m, mammals; p, plants. Vertical lines are as in Fig. 2.

These results varied by disturbance type; birds constituted the group most sensitive to forest conversion into agriculture (active agriculture, abandoned agriculture and agroforestry systems), whereas plants constituted the group most sensitive to burned forests and shaded plantations (Fig. 3a and Supplementary Table 2). The effect size for arthropods (0.64, 0.52–0.78) when further differentiated into the three main taxonomic orders revealed some differences: Coleoptera was more sensitive to disturbance (1.01, 0.75–1.30) than were Hymenoptera (0.41, 0.11–0.69) and Lepidoptera (0.58, 0.28–0.89) (Supplementary Table 1). In general, our findings reflect a paucity of information about most of the world's tropical biota; more data are needed to understand the ecological mechanisms underlying the differing vulnerability of taxa to human disturbance²⁷.

The source literature we considered used various measures of biodiversity, which we broadly differentiated into five response metrics: abundance, community structure and function, demographics, forest structure, and richness (Methods, Fig. 2c and Supplementary Fig. 1a). Of these, abundance and richness were the most commonly reported metrics, together comprising over three-quarters of all pairwise comparisons. Richness (0.83, 0.72–0.95) was markedly more sensitive to human disturbance than abundance (0.19, 0.07–0.31) (Figs 2c and

3b and Supplementary Tables 2 and 3). This result accords with expectations, given observations of large increases in the abundance of generalist species following similarly large declines in richness in degraded tropical forests^{5,28}. Furthermore, our measure of richness was predictably conservative because it assessed both forest specialists and generalists; when restricted to forest specialists ($n = 70$ comparisons), the effect size for species richness increased to 1.16 (0.69–1.65) (Fig. 2c and Supplementary Table 1). Measures of forest species richness therefore could serve as a simple yet effective metric to assess the conservation value of tropical forests and the relative impacts of different patterns of human modification, particularly during the early stages of forest conversion when conservation actions are most urgently needed.

We identified 12 general forest disturbance or conversion classes, and all but one of those with adequate sample sizes had effect sizes greater than 0.4 (Supplementary Table 1). In general, agricultural land-use classes (abandoned and active agricultural sites) had a much greater impact than agroforestry systems and plantations (both shaded and unshaded) (Fig. 2d). As the single exception, selectively logged forests (largely those affected by a single cutting cycle) had a much smaller, yet still positive, effect size of 0.11 (0.01–0.20). This is consistent with previous studies showing that selectively logged forests retain a high richness of forest taxa¹². Although these findings suggest that logged forests could contribute to biodiversity conservation, there are several caveats that need consideration: (i) if logged forest sites are adjacent to primary forests, spill-over effects might exaggerate the species richness of logged forests²² (acting as sink habitats); (ii) the proximity of logged forests to primary forests might also result in species extinction debts that are repaid over lengthy periods of time, beyond the timescale of the short-term studies that comprise most of our data set (83.6% had a time since disturbance of ≤ 12 yr); (iii) repeated logging might further exacerbate these biodiversity impacts; and (iv) the networks of forest roads created by logging operations might facilitate human immigration to forest frontiers and trigger associated increases in fires and forest conversion²⁹. As selective logging continues to expand across the tropics³⁰, understanding its long-term impacts and interactions with other forms of disturbance such as fire and invasive species⁵ will become increasingly important for the conservation of tropical biodiversity.

In contrast with the relatively benign selectively logged forests, secondary forests of varying ages had an intermediate effect size of 0.41 (0.28–0.54). It has been suggested recently that secondary forests can be an effective complement to primary forests in supporting tropical biodiversity, and should therefore represent a priority for conservation¹¹. Although the wide variety of secondary forests measured vary markedly in biodiversity value depending on forest age and land-use history, our meta-analysis demonstrates that secondary forests invariably have much lower biodiversity values than do remnant areas of relatively undisturbed primary forest (Supplementary Table 2). Although regenerating degraded areas can greatly increase the long-term persistence of biodiversity in severely modified landscapes⁶, our findings suggest that protecting remaining primary forests and restoring selectively logged forests are likely to offer the greatest conservation benefits for tropical biota.

We tested the relative importance of the above-mentioned ecological correlates in explaining the effect size. We used an information-theoretic approach to evaluate the performance of a candidate set of generalized linear models (Methods). After controlling for pseudo-replication from studies, the most parsimonious model in predicting the impact of anthropogenic forest disturbance on effect size was the null model (selected in 37.3% of 10,000 iterations), with the models 'Region' (23.1%) and 'Response metric' (14.4%) ranked second and third, respectively (Supplementary Table 5). This result also holds for a data set that includes only studies with natural vegetation as the surrounding habitat ($n = 1,557$), as well as for a smaller subset of data with information on time since disturbance and mean isolation

distance ($n = 630$; accounting for variation in colonization and succession effects²²) (Supplementary Fig. 2 and Supplementary Table 5). Our analysis of generalized linear models showed that the observed detrimental disturbance effects are essentially universal and that correlates such as region, taxonomic group, disturbance type and ecological measure have little impact on the effect size.

Our meta-analysis provides a global assessment of the relative conservation value of a broad range of human-modified tropical forests. Our results demonstrate that forest conversion and degradation consistently and greatly reduce biodiversity in tropical forest landscapes. As an exception, selective logging of forests has a much lower detrimental effect on measured biodiversity responses, implying that ecological restoration of such areas could help to alleviate threats to tropical biodiversity. Overall, however, we conclude that primary forests are irreplaceable for sustaining tropical biodiversity. Consequently, we strongly urge their protection by enhancing of enforcement in existing protected areas, expanding the current network of reserves and curbing international demand for forest commodities obtained at the expense of primary forests. Improving mechanisms for delivering and sustaining the social, financial and technical support necessary to achieve such goals continues to present one of the greatest challenges to tropical biodiversity conservation in the twenty-first century.

METHODS SUMMARY

Using Web of Science and BIOSIS, we searched for all relevant research articles published between 1975 and October 2010 that (i) included measures of biodiversity at multiple sites in both primary and disturbed tropical forests, (ii) indicated that the primary forests had little or no human disturbance and (iii) reported variance measures for biodiversity responses. From these studies, we compiled the biodiversity measures reported in both primary and disturbed forest sites and classified these measures using four variables: geographic region, taxonomic group, ecological response metric and disturbance type. For each paired biodiversity measure, we calculated the bias-corrected Hedges' g^* , the difference between primary and disturbed group means standardized by the pooled standard deviation. We then calculated the average effect size using the random-effects model, where effect sizes of individual comparisons are weighted by the inverse of within-study variance plus between-study variance²¹. We repeated this procedure after resampling the effect size calculations using 10,000 bootstrap samples (with replacement), from which we generated 95% confidence intervals. We calculated the effect size for the entire data set, for each subgroup of the four variables (region, taxon, response metric and disturbance type) and for each of the six two-level combinations of the four variables (for example disturbance type \times region). We repeated the above calculations for a subset of the data set with natural surrounding habitat, to account for the influence of this habitat. We also tested the effect sizes for possible publication bias. Following ref. 15, we performed an information-theoretic evaluation of a candidate set of generalized linear models to examine the influence of a set of proposed factors on the ecological responses tabulated. The generalized linear models related the Hedges' g^* effect size to the categorical predictor variables region, taxonomic group, metric and disturbance type in the 15 possible variable combinations.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 May; accepted 8 August 2011.

Published online 14 September 2011.

1. Dirzo, R. & Raven, P. H. Global state of biodiversity and loss. *Annu. Rev. Environ. Resour.* **28**, 137–167 (2003).
2. Sala, O. E. *et al.* Global biodiversity scenarios for the year 2100. *Science* **287**, 1770–1774 (2000).
3. Foley, J. A. *et al.* Global consequences of land use. *Science* **309**, 570–574 (2005).
4. Laurance, W. F. Have we overstated the tropical biodiversity crisis? *Trends Ecol. Evol.* **22**, 65–70 (2007).
5. Gardner, T. A. *et al.* Prospects for tropical forest biodiversity in a human-modified world. *Ecol. Lett.* **12**, 561–582 (2009).

6. Chazdon, R. L. Beyond deforestation: restoring forests and ecosystem services on degraded lands. *Science* **320**, 1458–1460 (2008).
7. Wright, S. J. Tropical forests in a changing environment. *Trends Ecol. Evol.* **20**, 553–560 (2005).
8. Barlow, J. *et al.* Quantifying the biodiversity value of tropical primary, secondary, and plantation forests. *Proc. Natl Acad. Sci. USA* **104**, 18555–18560 (2007).
9. Stork, N. E. *et al.* Vulnerability and resilience of tropical forest species to land-use change. *Conserv. Biol.* **23**, 1438–1447 (2009).
10. Gardner, T. A., Barlow, J., Sodhi, N. S. & Peres, C. A. A multi-region assessment of tropical forest biodiversity in a human-modified world. *Biol. Conserv.* **143**, 2293–2300 (2010).
11. Dent, D. H. & Wright, S. J. The future of tropical species in secondary forests: a quantitative review. *Biol. Conserv.* **142**, 2833–2843 (2009).
12. Edwards, D. P. *et al.* Degraded lands worth protecting: the biological importance of Southeast Asia's repeatedly logged forests. *Proc. R. Soc. B* **278**, 82–90 (2011).
13. Hughes, J. B., Daily, G. C. & Ehrlich, P. R. Conservation of tropical forest birds in countryside habitats. *Ecol. Lett.* **5**, 121–129 (2002).
14. Horner-Devine, M. C., Daily, G. C., Ehrlich, P. R. & Boggs, C. L. Countryside biogeography of tropical butterflies. *Conserv. Biol.* **17**, 168–177 (2003).
15. Sodhi, N. S., Lee, T. M., Koh, L. P. & Brook, B. W. A meta-analysis of the impact of anthropogenic forest disturbance on Southeast Asia's biotas. *Biotropica* **41**, 103–109 (2009).
16. Walpole, M. *et al.* Tracking progress toward the 2010 biodiversity target and beyond. *Science* **325**, 1503–1504 (2009).
17. Harvey, C. A., Dickson, B. & Kormos, C. Opportunities for achieving biodiversity conservation through REDD. *Conserv. Lett.* **3**, 53–61 (2010).
18. Strassburg, B. B. N. *et al.* Global congruence of carbon storage and biodiversity in terrestrial ecosystems. *Conserv. Lett.* **3**, 98–105 (2010).
19. Koh, L. P. & Ghazoul, J. Spatially explicit scenario analysis for reconciling agricultural expansion, forest protection, and carbon conservation in Indonesia. *Proc. Natl Acad. Sci. USA* **107**, 11140–11144 (2010).
20. Pereira, H. M. *et al.* Scenarios for global biodiversity in the 21st century. *Science* **330**, 1496–1501 (2010).
21. Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. *Introduction to Meta-Analysis* (Wiley, 2009).
22. Prugh, L. R., Hodges, K. E., Sinclair, A. R. E. & Brashares, J. S. Effect of habitat area and isolation on fragmented animal populations. *Proc. Natl Acad. Sci. USA* **105**, 20770–20775 (2008).
23. Peres, C. A., Barlow, J. & Laurance, W. F. Detecting anthropogenic disturbance in tropical forests. *Trends Ecol. Evol.* **21**, 227–229 (2006).
24. Koh, L. P. & Wilcove, D. S. Is oil palm agriculture really destroying tropical biodiversity? *Conserv. Lett.* **1**, 60–64 (2008).
25. Sodhi, N. S., Koh, L. P., Brook, B. W. & Ng, P. K. L. Southeast Asian biodiversity: an impending disaster. *Trends Ecol. Evol.* **19**, 654–660 (2004).
26. Daily, G. C., Ceballos, G., Pacheco, J., Suzán, G. & Sánchez-Azofeifa, A. Countryside biogeography of Neotropical mammals: conservation opportunities in agricultural landscapes of Costa Rica. *Conserv. Biol.* **17**, 1814–1826 (2003).
27. Gardner, T. A., Barlow, J., Parry, L. T. W. & Peres, C. A. Predicting the uncertain future of tropical forest species in a data vacuum. *Biotropica* **39**, 25–30 (2007).
28. Terborgh, J. *et al.* Ecological meltdown in predator-free forest fragments. *Science* **294**, 1923–1926 (2001).
29. Laurance, W. F., Goosem, M. & Laurance, S. G. Impacts of roads and linear clearings on tropical forests. *Trends Ecol. Evol.* **24**, 659–669 (2009).
30. Asner, G. P., Rudel, T. K., Aide, T. M., Defries, R. & Emerson, R. A contemporary assessment of change in humid tropical forests. *Conserv. Biol.* **23**, 1386–1395 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the tropical conservation scientists whose efforts helped to make our study possible. L.G., T.M.L. and N.S.S. were supported by grant R-154-000-479-112 from the National University of Singapore and L.G. was also supported by the Singapore International Graduate Award. L.P.K. was supported by the Swiss National Science Foundation and the ETH North-South Centre. T.A.G. thanks the Natural Environmental Research Council (NE/F01614X/1), and T.A.G. and J.B. thank the Instituto Nacional de Ciência e Tecnologia – Biodiversidade e Uso da Terra na Amazônia (CNPq 574008/2008-0) for funding. We dedicate this paper to the memory of N.S.S., who died while the manuscript was being reviewed.

Author Contributions The design of this project was the result of discussions involving all authors. L.G. and T.M.L. compiled the database; L.G., T.M.L. and B.W.B. performed the analysis; L.G., T.M.L. and N.S.S. wrote the initial draft of the manuscript; and all authors contributed to the writing of the final version of the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.G. (lgibson@nus.edu.sg) and T.M.L. (tmlee@ucsd.edu).

METHODS

Data. We searched for all relevant research articles published between 1975 and October 2010 using Web of Science and BIOSIS with the search query (TS = [(bird* OR mammal* OR reptile* OR amphibia* OR arthropod* OR plants* OR lepidoptera* OR hymenoptera* OR arachnid* OR coleoptera* OR diptera* OR homoptera* OR isoptera*) AND (clear-cutting* OR log* OR deforestation* OR fire* OR agriculture conversion* OR disturbance* OR degradation* OR secondary forest* OR plantation* OR fragment*)]). From this list, we reviewed articles and retained those studies that (i) included measures of biodiversity at multiple sites in both primary and disturbed tropical forests, (ii) indicated that the primary forests had little or no human disturbance and (iii) reported variance measures for biodiversity responses. We defined primary forests as primary or old-growth forests that have never been clear-felled and have been impacted by little or no known recent human disturbance.

For each study, we recorded the biodiversity measures in both primary and disturbed forest sites. For those studies that reported results in figures only, we extracted results using DATATHIEF (<http://www.datathief.org>). The full data set is available in the online version of the paper. For each comparison, we recorded the region (Africa, Asia, Central America (including Mexico), South America) and broad taxonomic group (arthropods, birds, mammals, plants). Although arthropods span diverse groups with potentially differing responses to human impacts⁸, our sample included predominantly insects (Coleoptera, 29.2%; Hymenoptera, 22.9%; Lepidoptera, 22.6%) and we therefore treated it as a single group but reported differences between the three major insect orders represented. Mammals also comprised different groups, and we differentiated between bats (51.0%), large mammals (2.6%), primates (3.7%), small mammals (28.2%) and a miscellaneous group (14.4%).

We classified the biodiversity measure into five response metrics: abundance (for example density, capture frequency, occupancy estimates and biomass); community structure and function (for example abundance of different guilds (generalists, herb specialists and so on), proportion of trait states and individual weight); demographics (for example density of different age classes (adults/juveniles/saplings/seedlings), fruit/flower production and genetic measures); forest structure (for example canopy height/cover/openness, basal area, litter depth, diameter at breast height and other physical structural measurements, and density of trees of a given diameter at breast height); and richness (for example observed/estimated/rarefied richness, species density and genera/family richness). We omitted diversity indices ($n = 151$; for example Fisher's alpha, Shannon–Wiener, Simpson's and Margalef's) because they were usually secondary (derived) measures of abundance and/or richness and are not straightforward to interpret.

We recorded the disturbance type as specified by the authors of the source literature, which formed twelve distinct groups: abandoned agriculture, active agriculture, agroforestry, burned forests, clear-cut forests, disturbed/hunted forests, other extracted forests, pastures, plantations, secondary forests, selectively logged forests and shaded plantations. To avoid an inadequate treatment of forest fragmentation, which is an important topic, we necessarily excluded data on forest fragments. However, we recognize that remnant forest fragments, particularly large ones, in heavily human-modified ecosystems might be critical for biodiversity persistence.

In addition, and where available, we collected data on patch size, surrounding habitat type, isolation distance and time since disturbance^{15,22}. We categorized the predominant surrounding habitat of disturbed forests into five broad groups: natural vegetation (that is, primary and selectively logged forests), agriculture, disturbed forests, pastures and tree plantations. Using maps and/or geo-referenced locations from the source literature, we calculated isolation distance as the mean distance between disturbed sites and the nearest primary forest site to account for colonization effects for a smaller set of the data. We measured time since disturbance as the amount of time that had elapsed between the most recent form of disturbance and the time of study, as indicated by the authors of the source literature, to account for post-disturbance and time-lag effects. We excluded patch size or area information from our analysis largely as a result of ambiguity and extremely low sample size (22.6% of the comparisons provided this information for disturbed sites). We have already acknowledged the potential confounding effects of area in detail elsewhere¹⁵. **Meta-analysis.** For each comparison, we calculated Hedges' g , the difference between primary and disturbed group means standardized using the pooled standard deviation of the two groups²¹, defined as:

$$g = \frac{x_{\text{primary}} - x_{\text{disturbed}}}{SD_{\text{pooled}}}$$

where

$$SD_{\text{pooled}} = \sqrt{\frac{(n_{\text{primary}} - 1)SD_{\text{primary}}^2 + (n_{\text{disturbed}} - 1)SD_{\text{disturbed}}^2}{n_{\text{primary}} + n_{\text{disturbed}} - 2}}$$

Because Hedges' g is a biased estimator of population effect size, we used the conversion factor J to compute a bias-corrected metric, g^* (ref. 21), defined as $g^* = Jg$, where

$$J = 1 - \frac{3}{4(n_{\text{primary}} + n_{\text{disturbed}} - 2) - 1}$$

We then calculated the average effect size using the random-effects model, where effect sizes of individual comparisons are weighted by the inverse of within-study variance plus between-study variance²¹. For individual comparisons, we defined the effect size as positive for comparisons where the biodiversity value was higher in primary forest (such that a positive effect size indicates a more detrimental impact by the disturbance type). For a small subset of comparisons where the expected value would be lower in primary forest ($n = 180$, 8.1% of all pairwise comparisons; for example measures of saplings/seedlings/juveniles, early/mid-successional species, non-forest/open-forest species, common/generalist/visitor species, trees of diameter at breast height <10 cm, dead/new trees and mortality/recruitment rates), we defined the effect size as negative for comparisons where the biodiversity value was higher in primary forest. As our results might be affected by the selection of comparisons with an opposite expectation of the direction of the effect, we repeated the procedure after omitting those comparisons. This led to an effect size of 0.45 (0.38–0.52), within the error of the effect size for the full data set, suggesting that our expectation did not affect the results (Supplementary Table 1).

We calculated the effect size for the entire data set, for each subgroup of the four variables (region, taxon, response metric and disturbance type) and for each of the six two-level combinations of the four variables (for example disturbance type \times region) (Fig. 3, Supplementary Fig. 1 and Supplementary Tables 2–4). For all combinations, we repeated this procedure after resampling the random-model effect size calculations using 10,000 bootstrap samples (with replacement), from which we generated 95% confidence intervals³¹. To address potential spatial and temporal autocorrelation from studies that included several comparisons (for example multiple measurements of the same taxa, measurements of multiple taxa and measurements of multiple disturbance types), we repeated this procedure after resampling one comparison per study, again using 10,000 bootstrap samples (Supplementary Table 1). However, some autocorrelation (largely only spatial) remains because several studies were situated in the same site (Fig. 1), although it is probably not as pronounced as above. To account for the potential influence of the surrounding habitat, we repeated the above calculations for a subset of the data set with natural surrounding habitat (70.1% of data) (Supplementary Table 1).

We tested for publication bias using two methods to assess whether our calculated effect sizes were affected by the possible absence of studies not published owing to a failure to detect differences²¹. First, we visually examined a funnel plot of effect size plotted against standard error to assess the symmetry of study precision around effect size (Supplementary Fig. 3). The relatively symmetrical funnel plot suggests there is no relationship between effect size and study size, and that those studies with small (or negative) effect sizes do not have a lower probability of being published. Second, we sorted the data set by precision, from comparisons with small standard errors to those with large standard errors, and examined the change in cumulative effect size with the addition of the most imprecise studies (Supplementary Fig. 4). Although the addition of the most imprecise third of comparisons (those with the largest standard errors) does cause the cumulative effect size to increase, the effect size remains positive and does not overlap with zero at any point after the first 163 comparisons. We conclude that the impact of publication bias in our study is slight²¹.

Generalized linear models. Following ref. 15, we performed an information-theoretic evaluation of a candidate set of generalized linear models (GLMs) to examine the influence of a set of hypothesized factors on the ecological responses tabulated. The GLM related the Hedges' g^* effect size to the categorical predictor variables region, taxonomic group, metric and disturbance type in the 15 possible variable combinations (Supplementary Table 5). We also evaluated the null (intercept-only) model, in which only a mean effect size is estimated (that is, no correlates). As with the meta-analysis, we accounted for pseudoreplication by selecting a random subset of the full data set, such that only one observation from each study was fitted using GLMs, and repeating the fitting procedure a total of 10,000 times. Model comparisons and subsequent inference (using relative weights of evidence) were based on the small-sample-size-corrected Akaike's information criterion (AIC_c ; ref. 32), whereby a measure of Kullback–Leibler information loss (a fundamental conceptual measure of the relative distance of a given model from full reality, assumed to be represented in the model set) is derived and used as an objective basis for ranking the bias-corrected likelihood of models in an a-priori candidate set (thereby yielding an implicit estimate of model parsimony). The highest-ranked models according to AIC_c are those that

explain the most substantial proportion of variance in the data yet exclude unnecessary parameters that cannot be justified for inference on the basis of the data³³. For the randomized GLM fits, we calculated the proportion of times each model was selected as the top-ranked model (π_i), on the basis of AIC_c . We used the per cent deviance explained to represent the structural goodness of fit of each model, with the 95% confidence interval of the per cent deviance explained estimated as the 2.5 and 97.5 percentiles of the 10,000 sample fits. We repeated the above analysis using only data with natural surrounding habitat, and using isolation distance and time since disturbance as additional predictor variables, thus increasing the possible variable combinations to 64 (including the null

model) (Supplementary Table 5). All statistical analyses and figures were made using the program R, version 2.11.1 (ref. 34).

31. Efron, B. & Tibshirani, R. Statistical data analysis in the computer age. *Science* **253**, 390–395 (1991).
32. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* 49–97 (Springer, 2002).
33. Burnham, K. P. & Anderson, D. R. Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildl. Res.* **28**, 111–119 (2001).
34. R Development Core Team. *The R Project for Statistical Computing*, version 2.11.1 (<http://www.R-project.org>) (2011).

Strong contributors to network persistence are the most vulnerable to extinction

Serguei Saavedra^{1,2,3*}, Daniel B. Stouffer^{4,5*}, Brian Uzzi^{1,2} & Jordi Bascompte⁴

The architecture of mutualistic networks facilitates coexistence of individual participants by minimizing competition relative to facilitation^{1,2}. However, it is not known whether this benefit is received by each participant node in proportion to its overall contribution to network persistence. This issue is critical to understanding the trade-offs faced by individual nodes in a network^{3–5}. We address this question by applying a suite of structural and dynamic methods to an ensemble of flowering plant/insect pollinator networks. Here we report two main results. First, nodes contribute heterogeneously to the overall nested architecture of the network. From simulations, we confirm that the removal of a strong contributor tends to decrease overall network persistence more than the removal of a weak contributor. Second, strong contributors to collective persistence do not gain individual survival benefits but are in fact the nodes most vulnerable to extinction. We explore the generality of these results to other cooperative networks by analysing a 15-year time series of the interactions between designer and contractor firms in the New York City garment industry. As with the ecological networks, a firm's survival probability decreases as its individual nestedness contribution increases. Our results, therefore, introduce a new paradox into the study of the persistence of cooperative networks, and potentially address questions about the impact of invasive species in ecological systems and new competitors in economic systems.

Mutualistic interactions form the basis of many biological and human systems of cooperation and competition^{2,6–14}. Mutualistic networks are composed of mutually beneficial interactions between individual participants or nodes of two distinct sets, such as plant species and their pollinators¹⁵ or designers and their contractors¹¹. One pattern in particular—nestedness—appears ubiquitous in mutualistic networks from a variety of contexts^{1,2,16}. In a nested network, the interactions are organized such that specialists (for example, plants with few pollinators) interact with proper subsets of the nodes with whom generalists (for example, plants with many pollinators) interact¹⁶. This nested architecture has been shown to minimize competition between species and therefore allows the network to support greater biodiversity¹.

Although greater nestedness allows for the successful coexistence of more species, it is unclear how the decreased risk of extinction is distributed among the nodes in the network. Here we quantify whether node-level survival benefits are related to each node's 'contribution' to the nested architecture, defined as the degree to which the organization of their interactions increases overall nestedness. A positive relationship could create a positive feedback of benefits to those that most support nestedness and the community; a negative relationship, in contrast, could imply that some nodes stand to benefit from the contributions of others.

To answer these questions, we integrate structural and dynamic analyses of 20 ecological networks that describe mutually beneficial interactions between flowering plants and their insect pollinators across

diverse environmental and biotic conditions¹⁷. In these bipartite networks, nodes correspond to individual plant or pollinator species and links between nodes indicate that a pollinator species has been found empirically to pollinate a given plant species (Methods).

To measure the individual contribution to nestedness for each species or node, we develop a novel, node-level metric that quantifies how an individual's contribution to network nestedness compares to that expected at random (Fig. 1). The measure quantifies the degree to which the overall nestedness of the network compares with the value obtained when randomizing just the interactions of that particular node.

Mathematically, this is defined as $c_i = (N - \langle N_i^* \rangle) / \sigma_{N_i^*}$, where N is the observed nestedness of the network and $\langle N_i^* \rangle$ and $\sigma_{N_i^*}$ are the average and standard deviation of nestedness across an ensemble of random replicates within which the interactions of node i have been randomized (Methods). The greater the degree to which the interactions of node i are consistent with the network's overall nestedness, the stronger is this node's contribution c_i , and vice versa.

We calculate this measure for each species in each of the networks and observe that node contributions to the network architecture are

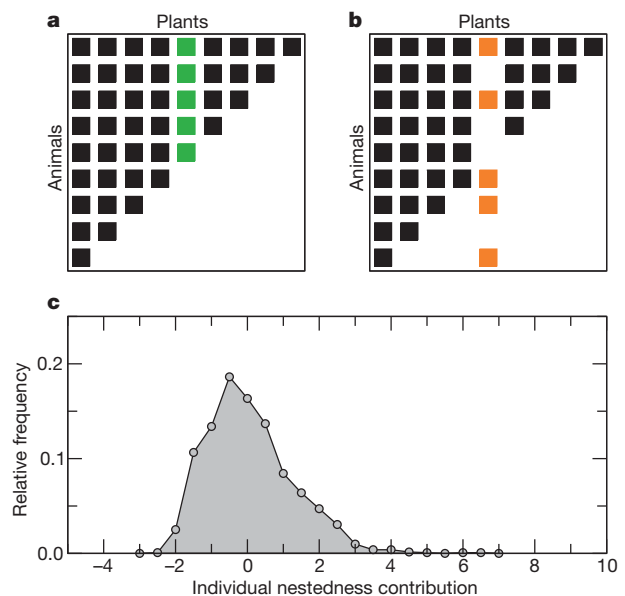


Figure 1 | Nodes contribute to the nested architecture of the network in distinct proportions. **a, b,** The individual nestedness contribution of a node (for example, the plant species whose interactions are highlighted in green in **a**) is defined as the degree to which the observed network nestedness compares to the value obtained when randomizing just the interactions of that particular node, highlighted in orange in **b**. **c,** The empirical distribution of individual nestedness contribution for all species in the 20 pollination networks studied here.

¹Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois 60208, USA. ²Kellogg School of Management, Northwestern University, Evanston, Illinois 60208, USA.

³Northwestern University Clinical and Translational Sciences Institute, Northwestern University, Chicago, Illinois 60611, USA. ⁴Integrative Ecology Group, Estación Biológica de Doñana, CSIC, Calle Américo Vespucio s/n, E-41092 Sevilla, Spain. ⁵School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand.

*These authors contributed equally to this work.

heterogeneously distributed across our empirical data set (Fig. 1c). Importantly, some species contribute comparatively little to the nested structure of the network whereas others contribute considerably more.

Having shown that species contribute to the overall network architecture in distinct proportions, we next explore the systemic consequences of species extinctions across the spectrum of contributions. This allows us to test whether or not a node's contribution in topological terms translates to a dynamic contribution in terms of network persistence. As there are no dynamic, empirical data with which to quantify network persistence in our ecological networks, we simulate species dynamics with a recently published model that is appropriate for mutualistic systems¹ (Methods). To measure the dynamic impact of a node on overall network persistence, we calculate the difference between the persistence of the network with and without removal of the focal node. Network persistence is measured as the fraction of initial species remaining at the end of the simulation. Here, we consider plant and pollinator species pooled together, while in Supplementary Information we calculate the persistence for each set independently.

We find that the more a node contributes to nestedness, the more likely it is that its loss is detrimental to the network's persistence (Fig. 2). Nestedness contribution therefore represents a key measure of the degree to which each node's interactions work for or against the long-term persistence of species in the mutualistic network.

Because the extinction of strong contributors has significant repercussions on network persistence, we proceed by estimating the vulnerability of these strong contributors to extinction. Specifically, we compare each node's nestedness contribution to its survival probability, where survival is determined by whether or not the node goes extinct before the dynamic simulations reach equilibrium.

Surprisingly, we find that nodes that contribute the most to the nestedness of the network—and its persistence—are the most likely to go extinct (Fig. 3). Indeed, individual nestedness contribution has a significant, negative correlation with survival probability (Methods). This conclusion is independent of whether all networks are analysed together ($P < 10^{-8}$ and $P < 10^{-15}$ for plants and pollinators, respectively) or each network is analysed separately. Specifically, the negative relationship between contribution and survival probability is significant for pollinators in 17 out of 20 networks and for plants in 20 out of 20 networks ($P < 10^{-4}$). Furthermore, our results capture the importance of nestedness contribution above and beyond the effect of the number of interactions per node (Methods). In general, the more a node contributes to the architecture of its network, the greater its

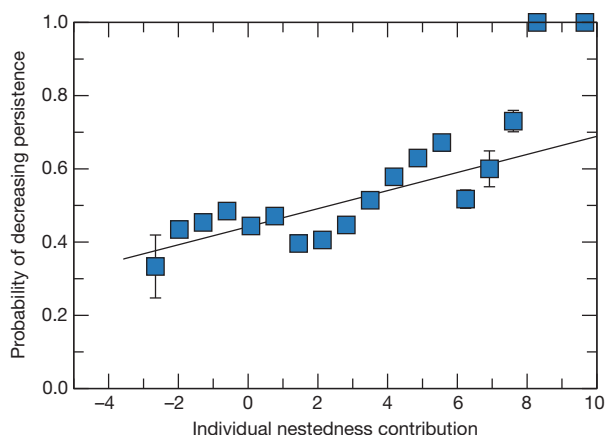


Figure 2 | The extinction of stronger contributors leads to a decrease in network persistence. We plot the probability that a node's removal causes a decrease in network persistence in the dynamic simulations as a function of that node's contribution to nestedness. All species in the 20 pollination networks are plotted together. Error bars, standard errors of the reported averages; in some cases they are smaller than the plotting symbols. Solid line, best-fit linear regression ($P < 0.005$).

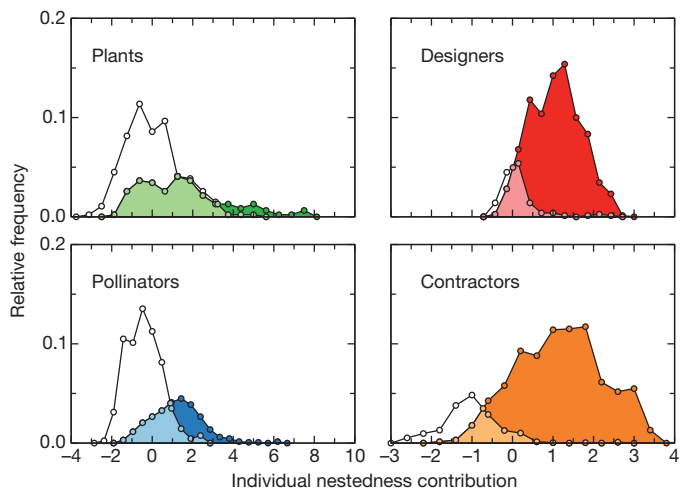


Figure 3 | Strong contributors to nestedness are the most vulnerable to extinction. We plot histograms of individual nestedness contribution for nodes that survive, with the area under the curve shaded in white, and for those that do not, with the area under the curve shaded in colour. In each case, the lighter colour indicates overlap between the two distributions. Nodes from all networks are pooled together for each class of node studied.

probability of extinction. These results imply that some species can benefit from others by participating in mutualistic interactions that differ from those that would maximize network persistence.

To explore the generality of these results to other types of cooperative networks, we build on previous work that illustrates commonalities between ecological and human systems^{2,9,11}. Specifically, we apply our node-contribution measure to the network formed by the cooperative interactions between designers and their contractors in the New York City garment industry across a 15-year observation period¹⁸. This industry is characterized by a competitive and dynamic environment where resource exchanges among firms and survival depends on collaborative links between firms^{11,18}. Though earlier studies have attributed the failure of firms to their lack of adaptation to new business networks^{19–21}, none have unambiguously linked firm survival to the network architecture. Here, nodes correspond to an individual designer or contractor firm, and links between nodes indicate that a designer exchanged money for the contractor's production services. We quantify a firm's survival by comparing whether or not the company was still operating at the end of the observation period.

Importantly, the negative relationship that we observed in ecological networks—between a node's contribution and its survival—also holds for the firms in this socio-economic network. This conclusion is obtained both when looking at the companies that went out of business by the end of the 15-year period ($P < 10^{-11}$ and $P < 10^{-12}$ for designers and contractors, respectively) and when tracking the yearly dynamics across the data set. In 14 out of 15 year-to-year intervals, the relationship between individual nestedness contribution and node survival is significant and negative for both contractors and designers ($P < 10^{-4}$). Just as for nodes in the ecological networks, the nodes that contribute the most to the nestedness of the network are the most likely to have failed (Fig. 3). This analysis suggests that our results may apply generally across different types of cooperative networks, whether they are shaped as a consequence of evolution or conscious decisions.

The individual nestedness contribution that we define here provides a means to estimate the expected survival of participants in cooperative systems solely from knowledge of the network structure. Moreover, we have revealed a paradox of nestedness: although strong contributors to nestedness are more important for the persistence of the entire network, they are also more prone to extinction compared to those nodes that contribute proportionally less. Although there is no clear explanation for this result, one could speculate that nodes that conform to the

architectural expectation of a nested network are subject to greater constraints than nodes that interact freely. If these linkage constraints have a cost in terms of fitness (for example, they lead to a pollinator species interacting with a less rewarding plant species), they could ultimately translate to a higher probability of extinction. Our study therefore raises new questions about the origins of nodes that make strong contributions to the collective good and nodes that appear to improve their own survival at the expense of others^{4,5,14}. In ecology, our results could inform a quantitative assessment of the likely persistence of an invasive species in the network and its effects on the overall welfare of the community. In socio-economic systems, our results could be used to identify those companies and economic sectors that undermine stable, long-term economic prosperity, as well as to develop interventions that take into account collective interests.

METHODS SUMMARY

Our ecological data set contains the 20 largest plant–pollinator mutualistic networks provided in the Supplementary Information of ref. 17; details about the networks and original sources can be found therein. Additionally, we analyse the network between designer and contractor firms in the New York City garment industry¹¹. Details about this temporal network can be found in the Supplementary Information of ref. 11.

Nestedness N is quantified using the measure proposed in ref. 22. In calculating nestedness contributions, the interactions of a node were randomized according to the null model specified in ref. 16; we used 1,000 random replicates. Note that all results presented in the present Letter hold both for alternative measures of nestedness and alternative null models (Supplementary Methods; Supplementary Figs 1–6).

To explore network dynamics, we employed the mutualistic model defined in ref. 1. This model is based on a system of differential equations describing the dynamics of P plant species and A animal species as a function of their intrinsic growth rates, interspecific competition, and mutualistic effects of one set on another.

We quantified the relationship between nestedness contribution c_i and node's survival probability s_i via a logistic regression. To ensure that our analysis is capturing a significant pattern over and above a node's degree (that is, its number of interactions), two additional terms in the regression controlled for the degree of the node and the potential interaction between degree and contribution to nestedness.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 April; accepted 5 August 2011.

Published online 14 September 2011.

1. Bastolla, U. *et al.* The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* **458**, 1018–1020 (2009).
2. Sugihara, G. & Ye, H. Cooperative network dynamics. *Nature* **458**, 979–980 (2009).
3. Rankin, D. J., Bargum, K. & Kokko, H. The tragedy of the commons in evolutionary biology. *Trends Ecol. Evol.* **22**, 643–651 (2007).
4. Leigh, E. G. How does selection reconcile individual advantage with the good of the group? *Proc. Natl Acad. Sci. USA* **74**, 4542–4546 (1977).

5. Falster, D. S. & Westoby, M. Plant height and evolutionary games. *Trends Ecol. Evol.* **18**, 337–343 (2003).
6. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
7. Gintis, H., Bowles, S., Boyd, R. T. & Fehr, E. (eds) *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life* (MIT Press, 2005).
8. Bascompte, J. Disentangling the web of life. *Science* **325**, 416–419 (2009).
9. May, R. M., Levin, S. A. & Sugihara, G. Complex systems: ecology for bankers. *Nature* **451**, 893–895 (2009).
10. Haldane, A. G. & May, R. M. Systemic risk in banking ecosystems. *Nature* **469**, 351–355 (2011).
11. Saavedra, S., Reed-Tsochas, F. & Uzzi, B. A simple model of bipartite cooperation for ecological and organisational networks. *Nature* **457**, 463–466 (2009).
12. Saavedra, S., Powers, S., McCotter, T., Porter, M. A. & Mucha, P. J. Mutually-antagonistic interactions in baseball networks. *Physica A* **389**, 1131–1141 (2010).
13. Fowler, J. H. & Christakis, N. Cooperative behavior cascades in human social networks. *Proc. Natl Acad. Sci. USA* **107**, 5334–5338 (2010).
14. Weyl, E. G., Frederickson, M. E., Yu, D. W. & Pierce, N. E. Economic contract theory tests models of mutualism. *Proc. Natl Acad. Sci. USA* **107**, 15712–15716 (2010).
15. Bascompte, J. & Jordano, P. The structure of plant–animal mutualistic networks: the architecture of biodiversity. *Annu. Rev. Ecol. Syst.* **38**, 567–593 (2007).
16. Bascompte, J., Jordano, P., Melián, C. J. & Olesen, J. M. The nested assembly of plant–animal mutualistic networks. *Proc. Natl Acad. Sci. USA* **100**, 9383–9387 (2003).
17. Rezendes, E. L., Lavabre, J. E., Guimarães, P. R., Jordano, P. & Bascompte, J. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* **448**, 925–928 (2007).
18. Saavedra, S., Reed-Tsochas, F. & Uzzi, B. Asymmetric disassembly and robustness in declining networks. *Proc. Natl Acad. Sci. USA* **10**, 16466–16471 (2009).
19. De Toni, A. & Nassimbeni, G. Supply networks: genesis, stability and logistics implications. A comparative analysis of two districts. *Omega* **23**, 403–418 (1995).
20. Uzzi, B. The sources and consequences of embeddedness for the economic performance of organizations: the network effect. *Am. Sociol. Rev.* **61**, 674–698 (1996).
21. Doeringer, J. & Crean, S. Can fast fashion save the US apparel industry? *Socioecon. Rev.* **4**, 353–377 (1996).
22. Almeida-Neto, M., Guimarães, P., Guimarães, P. R. Jr, Loyola, R. D. & Ulrich, W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117**, 1227–1239 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Funding was provided by the Kellogg School of Management, Northwestern University, the Northwestern University Institute on Complex Systems (NICO; to S.S. and B.U.), NUCATS grant UL1RR025741 (to S.S.), a CSIC-JAE postdoctoral fellowship (to D.B.S.), the Army Research Laboratory (under cooperative agreement W911NF-09-2-0053 to B.U.), and the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) through an Advanced Grant (grant agreement 268543 to J.B.). Figures were generated with PyGrace (<http://pygrace.sourceforge.net>).

Author Contributions S.S. and D.B.S. analysed the data and performed the simulations. S.S., D.B.S., B.U. and J.B. designed the study and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.B. (bascompte@ebd.csic.es).

METHODS

Data sets. Here we analyse two types of empirical data. First, we analyse a data set containing the 20 largest plant–pollinator mutualistic networks provided in the Supplementary Information of ref. 17. Further details about the network size, species composition and geographic location can be found in ref. 17, together with the actual network and original source. Additionally, we analyse the socio-economic network first described in ref. 11. This network contains approximately 700,000 commercial interactions between manufacturer and contractor firms in the New York City garment industry between January 1985 and December 2003. From this data set, we can generate yearly snapshots of the bipartite network. Additional details can be found in the Supplementary Information of ref. 11.

Nestedness measure. We quantify nestedness using NODF, the measure recently introduced in ref. 22. This measure reduces potential bias introduced by network size and shape compared with alternative measures. The overall nestedness of the network N is defined mathematically as:

$$N = \frac{\sum_{i < j}^P M_{ij} + \sum_{i < j}^A M_{ij}}{\left[\frac{P(P-1)}{2} \right] + \left[\frac{A(A-1)}{2} \right]}$$

where the first sum is across all pairs of plant species, the second sum is across all pairs of animal species, and P and A are the total number of plant species and animal species, respectively. For every pair of nodes i and j , $M_{ij} = 0$ if $k_i = k_j$, and $M_{ij} = n_{ij}/\min(k_i, k_j)$ otherwise. Here, k is a node's number of interactions; n_{ij} is the number of interactions in common between nodes i and j ; and $\min(k_i, k_j)$ refers to the minimum of the two values k_i and k_j .

This nestedness metric takes values in the interval $N \in [0, 1]$, where 1 designates a perfectly nested network and 0 indicates a network with no nestedness. Alternative measures of nestedness are highly and significantly correlated to this one (Supplementary Methods; Supplementary Fig. 1). Furthermore, all results presented in this Letter hold for such alternative measures of nestedness (Supplementary Figs 2 and 3).

Null model. To randomize interactions, we use the null model outlined in ref. 16. Under the specifications of this null model, the interactions are assigned according to the rule:

$$p_{ij} = \frac{1}{2} \left(\frac{k_i}{P} + \frac{k_j}{A} \right)$$

where p_{ij} is the probability of an interaction between node i (of set A) and node j (of set P) and A and P are the total number of animal and plant species in sets A and P ,

respectively. In the socio-economic networks, the model is specified in the same fashion except with designers and contractors in the place of animals and plants. We use 1,000 replicates. Note that all results presented in this Letter hold for alternative null models (Supplementary Methods; Supplementary Figs 4, 5, and 6).

Dynamic model for ecological mutualistic networks. To simulate interspecies dynamics in the ecological networks, we run the dynamic model of ref. 1 using each of the real networks as the skeleton of the model. The real networks specify the number of plant species, the number of animal species, and who interacts with whom. In the dynamic model, the change in abundance over time for a plant species i follows:

$$\frac{dS_i^{(P)}}{dt} = \alpha_i^{(P)} S_i^{(P)} - \sum_{j \in P} \beta_{ij}^{(P)} S_i^{(P)} S_j^{(P)} + \sum_{k \in A} \frac{\gamma_{ik}^{(P)} S_i^{(P)} S_k^{(A)}}{1 + h^{(P)} \sum_{l \in A} \gamma_{il}^{(P)} S_l^{(A)}}.$$

The same equations for animal species can be written in a symmetric form interchanging the indices (P) and (A) .

To fully specify the remainder of the dynamic model, we use the following parameter values for all plant species (P) and animal species (A): intrinsic growth rates α_i are drawn uniformly from the interval $[0.85, 1.1]$; the competitive interactions β_{ii} and β_{ij} are drawn uniformly from the intervals $[0.99, 1.01]$ and $[0.22, 0.24]$, respectively; the mutualistic interactions γ_{ij} , encapsulating the *per capita* effect of animal j on plant i , are drawn uniformly from the interval $[0.19, 0.21]$; the handling time h is set to 0.1.

Simulations are performed by integrating the system of ordinary differential equations using a fourth-order Runge-Kutta method with small integration steps. All initial abundance densities S_i are drawn uniformly from the interval $(0, 1]$. Species are considered to have gone extinct when their abundance density S_i is lower than 10^{-30} . All results are robust to changes in parameter values (growth rates $\alpha_i \in [0, 2]$, competitive interactions $\beta_{ij} \in [0, 1]$, and mutualistic interactions $\gamma_{ij} \in [0, 1]$) as well as changing the functional responses of the above model from Holling type II to Holling type III.

Relationship between nestedness contribution and node survival. We quantify the relationship between nestedness contribution c_i and node's survival probability s_i by using a logistic regression with the form $\text{logit}(s_i) = \alpha + \beta c_i$. Survival was coded as 0 and 1 for non-surviving and surviving nodes, respectively. To ensure that our analysis is capturing a significant pattern above and beyond other network attributes, we perform the same analysis but also include terms for node degree k (that is, number of interactions) and the potential interaction term for degree and contribution. This extended model takes the form $\text{logit}(s_i) = \alpha + \beta c_i + \gamma k_i + \delta c_i k_i$.

CTCF-binding elements mediate control of V(D)J recombination

Chunguang Guo^{1*}, Hye Suk Yoon^{1*}, Andrew Franklin^{1*}, Suvi Jain¹, Anja Ebert², Hwei-Ling Cheng¹, Erica Hansen¹, Orion Despo¹, Claudia Bossen³, Christian Vettermann⁴, Jamie G. Bates⁴, Nicholas Richards¹, Darienne Myers¹, Harin Patel¹, Michael Gallagher¹, Mark S. Schlissel⁴, Cornelis Murre³, Meinrad Busslinger², Cosmas C. Giallourakis^{1,5} & Frederick W. Alt¹

Immunoglobulin heavy chain (IgH) variable region exons are assembled from V_H, D and J_H gene segments in developing B lymphocytes. Within the 2.7-megabase mouse *Igh* locus, V(D)J recombination is regulated to ensure specific and diverse antibody repertoires. Here we report in mice a key *Igh* V(D)J recombination regulatory region, termed intergenic control region 1 (IGCR1), which lies between the V_H and D clusters. Functionally, IGCR1 uses CTCF looping/insulator factor-binding elements and, correspondingly, mediates *Igh* loops containing distant enhancers. IGCR1 promotes normal B-cell development and balances antibody repertoires by inhibiting transcription and rearrangement of D_H-proximal V_H gene segments and promoting rearrangement of distal V_H segments. IGCR1 maintains ordered and lineage-specific V_H(D)J_H recombination by suppressing V_H joining to D segments not joined to J_H segments, and V_H to DJ_H joins in thymocytes, respectively. IGCR1 is also required for feedback regulation and allelic exclusion of proximal V_H-to-DJ_H recombination. Our studies elucidate a long-sought *Igh* V(D)J recombination control region and indicate a new role for the generally expressed CTCF protein.

The variable region exons of IgH, Ig light (IgL) and T-cell receptor genes are assembled during B- or T-cell development from variable (V), diversity (D) and joining (J) gene segments¹. The V(D)J recombination reaction is initiated by RAG endonuclease¹. RAG cleaves only paired gene segments flanked, respectively, by complementary recombination signals (RSs) referred to as 12RSs and 23RSs, a restriction referred to as the 12/23 rule¹. The cleaved segments are then fused via classical non-homologous end-joining (C-NHEJ)². The mouse *Igh* locus contains hundreds of V_H gene segments within a several-megabase (Mb) region, followed downstream by a 100 kilobase (kb) 'intergenic' region separating the most downstream V_H (generally referred to as V_{H81X}, but formally denoted V_{H7183.a2.3}; NCBI accession number AJ851868)³ from D_{FL16.1}, the first of 13 clustered D_H segments. The most downstream D (D_{Q52}) lies upstream of 4 J_H segments (J_{H1}–J_{H4})⁴. V_H and J_H gene segments are flanked by 23RSs and D segments are flanked on both sides by 12RSs, ensuring that V_H(D)J_H assembly involves joining V_H and J_H segments to the upstream and downstream sides of a D_H segment, respectively⁴. The *Igh* constant region (C_H) exons lie in the 200-kb region downstream of the J_H segments; RNA splicing fuses productively assembled V_H(D)J_H and C_H exons during *Igh* messenger RNA formation.

Igh V(D)J recombination in developing B cells is regulated to be highly ordered and stage specific; thus, D_H-to-J_H joining developmentally occurs first on both alleles in pre-progenitor (pro)-B cells followed by appendage of a V_H to a DJ_H complex in pro-B cells^{4–6}. Direct joining of a V_H to an un-rearranged D_H does not occur, even though theoretically permitted by the 12/23 rule^{3,7}. The V_H-to-DJ_H joining step is also regulated to achieve lineage specificity; thus, although developing T cells generate DJ_H joins, they do not form complete V_H(D)J_H exons^{7,8}. At the pro-B stage, V(D)J recombination is regulated in the context of allelic exclusion, with a signal from a productive (that is, μ IgH protein-encoding) V_H(D)J_H rearrangement

inhibiting V_H-to-DJ_H joining on the other *Igh* allele, if it is in the DJ_H configuration⁵. Expression of the μ chain also signals development to the precursor (pre)-B cell stage and *Igl* V(D)J recombination⁹. To generate such signals in pro-B cells, μ IgH chains must pair with surrogate IgL chains¹⁰. Subsequently, μ chains must pair with IgL chains in pre-B cells to mediate the pre-B-to-IgM⁺ B-cell transition. Lastly, *Igh* V(D)J recombination is regulated to ensure utilization of V_H segments across the large V_H locus. However, proximal V_H segments, particularly V_{H81X}, are rearranged more frequently than distal V_H segments, leading to over-representation in primary V_H(D)J_H repertoires³. Repertoire normalization for distal V_H segments in mature B cells relies on cellular selection^{3,11}, promoted, in part, by the inability of certain proximal V_H segments, including V_{H81X}, to pair with surrogate IgL chains and IgL chains^{12,13}.

V(D)J recombination at all antigen receptor loci is effected by the common V(D)J recombinase comprised of RAG and C-NHEJ components. Regulation of *Igh* V(D)J recombination in the context of order/stage, lineage and allelic exclusion is achieved via modulation of substrate V, D and J accessibility^{14,15}. Correlates of such accessibility include transcription of un-rearranged gene segments and certain DNA and histone modifications^{4,14–21}. *Igh* locus contraction and looping may also mediate higher-order regulation of V(D)J recombination, for example by bringing distant V_H segments into proximity with the DJ_H^{8,22–24}. Until now, *cis* elements that control order, lineage-specificity, allelic exclusion and/or differential V_H utilization have been elusive^{16,19}. The only known long-range *Igh* regulatory elements are a transcriptional enhancer (termed iE μ) in the intron between the J_H and C_H segments and a set of long-range enhancers (termed the 3' *Igh* regulatory region) downstream of the C_H segments^{18,25}. The iE μ transcriptional enhancer is required for efficient *Igh* V(D)J recombination, particularly V_H-to-DJ_H joining^{26–28}, although the mechanisms by which it influences this process are unknown¹⁸. Thus far, the 3' *Igh*

¹Howard Hughes Medical Institute, The Children's Hospital, The Immune Disease Institute; Department of Genetics, Harvard Medical School, 300 Longwood Avenue, Boston, Massachusetts 02115, USA.

²Research Institute of Molecular Pathology, Dr. Bohr-Gasse 7, A-1030 Vienna, Austria. ³Division of Biological Sciences, 0377, University of California, San Diego, La Jolla, California 92093, USA. ⁴Department of Molecular & Cell Biology, University of California, Berkeley, California 94720, USA. ⁵Gastrointestinal Unit, Massachusetts General Hospital, 55 Fruit Street, Boston, Massachusetts 02114, USA.

*These authors contributed equally to this work.

regulatory region has not been implicated in V(D)J recombination²⁵. As most critical aspects of *Igh* V(D)J recombination are regulated at the V_H-to-DJ_H step^{7,8}, relevant regulatory elements may reside in the 100 kb intergenic region separating the V_H and D_H segments (see Supplementary Discussion)^{7,16,17,19,29,30}.

Role in normal B-cell development

The region several kilobases upstream of D_{FL16.1} harbours chromatin modifications^{29,31,32} and two CTCF-binding elements (CBEs)^{29,31–33} suggestive of a potential regulatory region (Supplementary Fig. 1). CTCF is an 11-zinc-finger nuclear protein implicated in transcriptional insulation, chromatin boundary formation, transcriptional activation/repression and chromosome looping^{34–36}. There are several other potential *cis*-elements closely linked to these CBEs including potential PU.1³¹ and YY1-binding sites (using the JASPAR database). We refer to this cluster of factor-binding sites as IGCR1 (Fig. 1). To test for a role in *Igh* V(D)J recombination, we generated an IGCR1-deleted 129SV allele in which the 4.1-kb DNA fragment that contains

both the CBEs and other binding sites was deleted in the mouse germ line (Fig. 1a and Supplementary Fig. 2). To test for specific roles of the CBEs, we generated mice in which both were replaced with scrambled sequences that do not bind CTCF (Supplementary Figs 1 and 3). Mice heterozygous or homozygous for the IGCR1 deletion are referred to, respectively, as IGCR1^{+/-} and IGCR1^{-/-}, and mice heterozygous or homozygous for the dual CBE mutation are referred to, respectively, as IGCR1/CBE^{+/-} or IGCR1/CBE^{-/-}. Because generation of mutant alleles involved *loxP* insertion, we generated control lines heterozygous or homozygous for the *loxP* insertion referred to, respectively, as *loxP*^{+/-} or *loxP*^{+/+} (Fig. 1a). As wild-type, *loxP*^{+/-} and *loxP*^{+/+} mice gave essentially identical results, we refer to them collectively as ‘controls’. As a further control, we deleted an approximately 2-kb DNA fragment downstream of the D_H-proximal end of IGCR1 and found no obvious phenotype (Supplementary Fig. 10).

IGCR1/CBE^{+/-} or IGCR1/CBE^{-/-} mice had similar splenic IgM⁺ B-cell numbers as controls (Fig. 1b and Supplementary Fig. 5a). However, IGCR1/CBE^{+/-} and, more so, IGCR1/CBE^{-/-} mice had

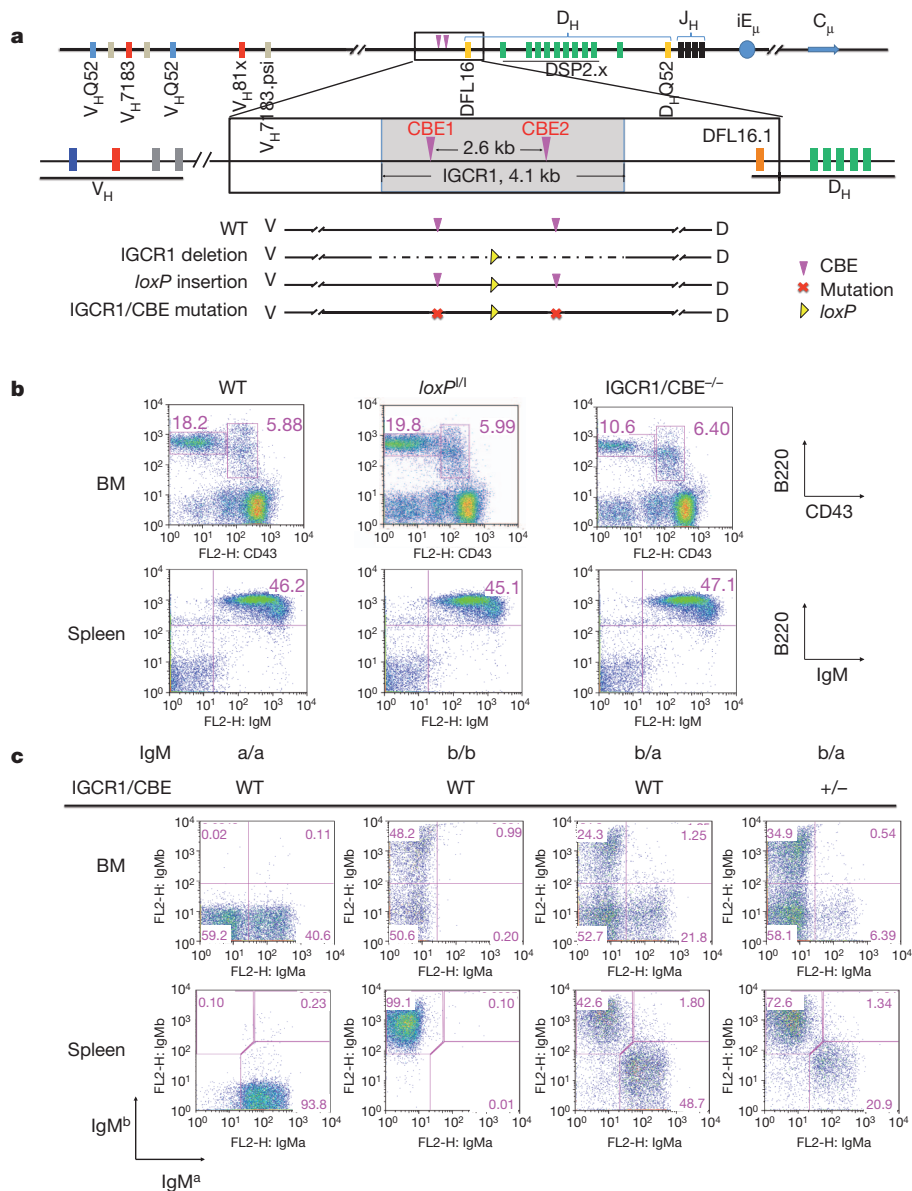


Figure 1 | Mutation of IGCR1 CBEs impairs B-cell development. **a**, Murine 129SV *Igh* locus (NCBI accession number: AJ851868) schematic showing the 4.1-kb IGCR1 region in wild type (WT) compared to IGCR1-deleted, *loxP*-inserted, or CBE-mutated configuration. **b**, Flow cytometry analysis of IgM⁺ bone marrow (BM) and IgM⁺ splenic B-cell populations in wild-type, *loxP*^{+/+}

and IGCR1/CBE^{-/-} mice. In bone marrow the B220^{int}CD43⁺ pro-B and B220⁺CD43⁻ pre-B cell populations are indicated. **c**, Expression of IgM^a and IgM^b allotypic markers in bone marrow and spleen from wild-type IgM^a/IgM^a (pure 129SV), wild-type IgM^b/IgM^b (pure C57BL/6), wild-type F1 (IgM^a/IgM^b) and heterozygous mutant IGCR1/CBE^{+/-} IgM^a/wild-type IgM^b mice.

a substantial diminution in bone marrow pre-B cell numbers (Fig. 1b and Supplementary Fig. 5b). As the pro-B-to-pre-B transition is signalled by a productive $V_H(D)J_H$ in pro-B cells, this developmental defect suggests an *Igh* V(D)J recombination defect. As a more sensitive test for the roles of IGCR1 in B-cell development, we bred 129SV IGCR1/CBE^{+/-} mice with C57BL/6 wild-type mice to generate F1 mice with a wild-type *Igm*^b allele and a CBE-mutated *Igm*^a allele and assayed B cells for surface IgM^a and IgM^b expression. Remarkably, whereas normal F1 mice, as expected, have roughly equal numbers of IgM^a- and IgM^b-expressing B cells (but not both due to *Igh* allelic exclusion), most IgM⁺ bone marrow and splenic B cells in F1 mice carrying the IGCR1 CBE-mutant *Igm*^a allele express IgM^b (Fig. 1c). Thus, mutation of the IGCR1 CBEs renders an *Igh* allele ineffective in supporting B-cell development when competing against a wild-type *Igh* allele. We found identical B-cell developmental defects in IGCR1^{+/-} and IGCR1^{-/-} mice (Supplementary Figs 4b, c and 5c, d).

Mediation of diverse *Igh* repertoires

We used a polymerase chain reaction (PCR) approach (Supplementary Fig. 6a) to assay for DJ_H and V_H(D)J_H rearrangements in purified control, IGCR1/CBE^{+/-}, IGCR1/CBE^{-/-}, IGCR1^{+/-} and IGCR1^{-/-} bone marrow pro-B and pre-B cells, and in splenic B cells. We assayed for rearrangements of the two most D_H-proximal V_H families (V_H7183 and V_HQ52) and the most distal V_H family (V_HJ558). *Igl* V_κ-to-J_κ joins were assayed as a stage-specific control and the mouse *Dlg5* gene as a loading control. Levels of DJ_H and V_κJ_κ rearrangements did not vary markedly among different populations or genotypes; thus, V(D)J recombination in general was not affected by the mutations (Fig. 2a and Supplementary Fig. 6). However, relative levels of proximal V_H7183DJ_H rearrangements were markedly increased and those of distal V_HJ558DJ_H rearrangements markedly reduced in IGCR1/CBE^{-/-} and IGCR1^{-/-} pro-B cells, with both being intermediate in IGCR1/CBE^{+/-} and IGCR1^{+/-} pro-B cells (Fig. 2a and Supplementary Fig. 6).

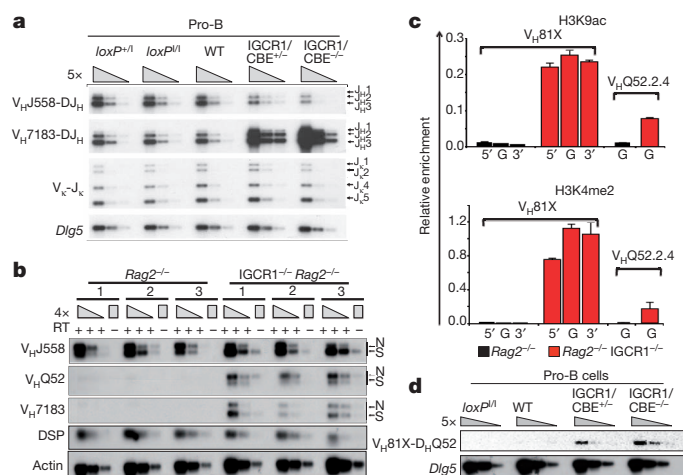


Figure 2 | IGCR1 mutations alter V_H usage, germline transcription and rearrangement order. **a**, PCR analyses of indicated V_H family rearrangements in pro-B cells from indicated mice compared to a *Dlg5* loading control. Results are typical of four experiments. Bands corresponding to rearrangements to various J_H segments are indicated on right. **b**, RT-PCR analysis of indicated germline V_H transcripts in three independent wild-type and IGCR1^{-/-} A-MuLV-virus-transformed *Rag2*^{-/-} pro-B-cell lines. N, nonspliced sense/antisense; S, spliced sense. **c**, ChIP-qPCR analyses of H3K4me2 and H3K9ac histone modifications at indicated V_H segments in 129SV *Rag2*^{-/-} (black) and *Rag2*^{-/-} IGCR1^{-/-} (red) A-MuLV-transformed pro-B lines. The 5' region (5'), gene body (G) and 3' region (3') of V_H81X and gene body (G) of V_HQ52.2.4 were analysed. Average values and standard deviations of three experiments with one line shown are representative of results from both. **d**, Semi-quantitative PCR analyses of direct V_H-to-D rearrangements in sorted pro-B cells from indicated mice. The PCR assays used for panels **a**, **b** and **d** are diagrammed in Supplementary Figs 6a, 7a and 8a.

Within the two proximal V_H families, V_H usage was even more skewed towards the most D-proximal members in IGCR1^{-/-} pro-B cells (Supplementary Fig. 6c). Together, these findings are consistent with IGCR1 mutations resulting in *cis*-acting increases and *cis*-acting decreases, respectively, in proximal and distal V_H rearrangement. Given that the proximal V_H segments contribute to a substantial fraction of V_H(D)J_H rearrangements (about 40%) in normal pro-B cells^{3,11}, increased V_H7183 joins in IGCR1/CBE^{+/-} and IGCR1^{+/-} pro-B cells indicates that the absolute level of V_H-to-DJ_H rearrangements on mutant alleles, although even more biased towards proximal V_H segments than normal, is not decreased. In the various IGCR1 mutant pre-B cells and splenic IgM⁺ B cells repertoire bias remained; although the extent was progressively moderated (Supplementary Fig. 6), probably due to cellular selection for V_H repertoire normalization.

Regulation of germline V_H transcription

To measure germline V_H transcripts, we generated *Rag2*-deficient Abelson murine leukaemia virus (A-MuLV)-transformed wild-type, IGCR1^{+/-} and IGCR1^{-/-} pro-B lines. *Rag2*-deficient lines have unrearranged *Igh* alleles; thus, any detected V_H transcripts are germline. RNA was assayed via reverse transcriptase PCR (RT-PCR) for V_H expression, using one primer from the V_H leader sequence and another from downstream of the RS (Supplementary Fig. 7a). On the basis of size, the PCR assay detects both unspliced germline V_H transcripts (sense or antisense) and slightly smaller, spliced sense germline V_H transcripts (Fig. 2b). *Rag2*^{-/-} pro-B lines had robust D_H transcripts and spliced and un-spliced V_HJ558 transcripts, but lacked readily detectable V_HQ52 or V_H7183 transcripts (Fig. 2b). However, *Rag2*^{-/-} IGCR1^{+/-} and, more so, *Rag2*^{-/-} IGCR1^{-/-} pro-B lines showed marked upregulation of spliced and unspliced V_HQ52 and V_H7183 transcripts with normal levels of V_HJ558 and D_H transcripts (Fig. 2b and Supplementary Fig. 7d). We even detected by northern blotting a ~3.5-kb V_H81X-hybridizing transcript in RNA from *Rag2*^{-/-} IGCR1^{-/-} lines, but not in wild-type *Rag2*^{-/-} lines (Supplementary Fig. 7f). Primary *Rag2*^{-/-} IGCR1^{-/-} pro-B cells also strongly upregulated germline V_H7183 transcripts (Supplementary Fig. 7e). Lastly, chromatin immunoprecipitation-sequencing (ChIP-seq) and chromatin immunoprecipitation-quantitative PCR (ChIP-qPCR) analyses revealed that deletion of IGCR1 led to a marked increase in active histone marks over V_H81X (V_H7183.a2.3) and the adjacent V_HQ52.a2.4 germline gene segments (Fig. 2c and Supplementary Fig. 7b, c). Thus, IGCR1 suppresses activation of germline V_H segments over distances of at least 100 kb.

Role in order and lineage specificity

We assayed for V_H81X-to-germline-DQ52 joins via PCR with a forward V_H81X-specific primer and a reverse primer from sequences between DQ52 and J_H1 (Supplementary Fig. 8). Whereas we did not detect direct V_H81X-to-DQ52 joins in control pro-B cells, we readily detected them in IGCR1/CBE^{+/-}, IGCR1/CBE^{-/-}, IGCR1^{+/-} and IGCR1^{-/-} pro-B cells (Fig. 2d and Supplementary Fig. 8). Sequences of 133 independent direct V_H7183DQ52 joins revealed that 120 involved V_H81X, 12 involved the downstream pseudo-V_H7183, and one involved the next V_H7183 upstream of V_H81X (Supplementary Table 2). Therefore, integrity of the IGCR1 CBEs is required for ordered *Igh* V(D)J recombination in pro-B cells, at least for proximal V_H segments.

To examine potential IGCR1 roles in lineage-specific *Igh* V(D)J recombination, we assayed for D-to-J_H, V_H-to-DJ_H and V_κ-to-J_κ rearrangements in DNA from CD4⁺CD8⁺ (double-positive) thymocytes from control and IGCR1/CBE^{+/-}, IGCR1/CBE^{-/-}, IGCR1^{+/-} and IGCR1^{-/-} mice (Fig. 3a and Supplementary Fig. 9). We detected DQ52J_H rearrangements in all mice (Fig. 3a and Supplementary Fig. 9). However, whereas there were no V_H(D)J_H rearrangements in controls, we readily detected V_H(D)J_H rearrangements of proximal V_H7183 and V_HQ52 segments, but not distal V_HJ558 segments, in mutant double-positive thymocytes (Fig. 3a and Supplementary Fig. 9). Lack of V_κJ_κ

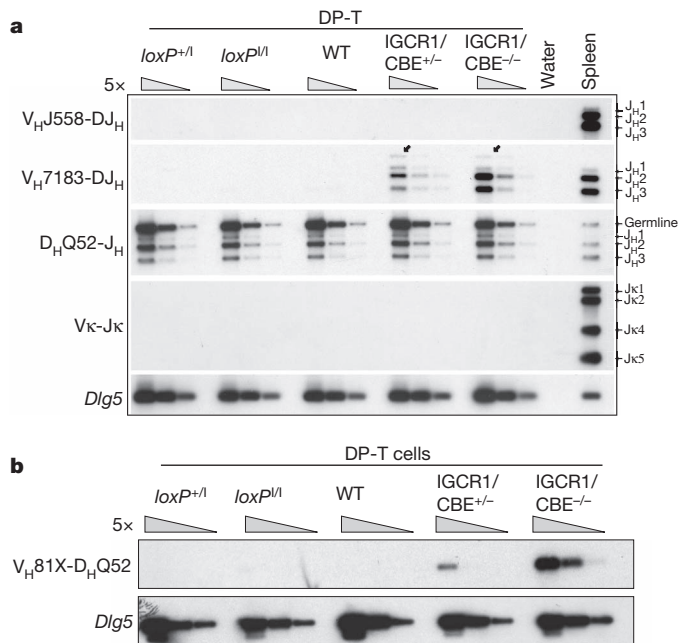


Figure 3 | IGCR1/CBE mutations lead to $V_H(D)J_H$ and V_HD rearrangements in thymocytes. **a**, PCR analyses of V_H family rearrangements in sorted double-positive thymocytes (DP-T) and total splenic B cells from indicated mice with *Dlg5* as a loading control. Bands corresponding to rearrangements to various J_H segments are indicated on right. *Igκ* rearrangement ($V_κJ_κ$) served as a control for B-cell contamination. **b**, Semi-quantitative PCR analyses of direct V_H -to-D rearrangements in sorted double-positive T cells (DP-T cells) from indicated mice. Assays are diagrammed in Supplementary Figs 6a and 8a.

rearrangements confirmed absence of B-cell contamination. Cloning and sequencing of V_{H7183} - and V_{H52} -to- DJ_H rearrangements from IGCR1^{-/-} double-positive thymocytes revealed predominant utilization of the most proximal V_H segments (V_{H81X} and $V_{H52.a2.4}$; Supplementary Tables 3 and 4). We also assayed for direct V_{H81X} -to-germline D_{Q52} joins in double-positive thymocytes (Fig. 3b and Supplementary Fig. 8). As expected, controls lacked detectable direct V_H -to-D joins; but such joins were readily apparent in mutant thymocytes (Fig. 3b and Supplementary Fig. 8). Nucleotide sequencing of 32 V_HD joins revealed 29 used V_{H81X} and the rest used the downstream pseudo- V_{H7183} (Supplementary Table 2). Thus, IGCR1 CBEs are required for lineage-specific *Igh* V_H -to- DJ_H recombination.

Role in proximal V_H feedback regulation

Surface staining of splenic B cells heterozygous for the IGCR1-deleted *Igm^a* allele and a wild-type *Igm^b* allele did not reveal allelic inclusion (Supplementary Fig. 4c). Likewise, no *Igm^a/Igm^b* double expressers were found in nearly 900 individual IGCR1^{+/-} F1 splenic B cells by

cytoplasmic staining (Supplementary Fig. 11a). Hybridoma analyses showed that about 60% of wild-type B cells had a productive $V_H(D)J_H$ on one allele and a DJ_H on the other (that is, $V_H(D)J_H^+/DJ_H^-$ configuration) and about 40% had $V_H(D)J_H$ rearrangements on both alleles (that is, $V_H(D)J_H^+/V_H(D)J_H^-$ configuration) (Fig. 4a). This 60/40 ratio reflects feedback regulation of V_H -to- DJ_H joining from productive rearrangements^{5,6}. In IGCR1^{+/-} B cells, this ratio inverted to 30/70, demonstrating that heterozygous IGCR1 deletion markedly increases B cells with $V_H(D)J_H$ joins on both alleles, despite allelic exclusion at the protein level. Analyses of 39 $V_H(D)J_H/V_H(D)J_H$ IGCR1^{+/-} B-cell hybridomas revealed that most had a $V_H(D)J_H^+$ that used a distal V_H and a $V_H(D)J_H^-$ that used V_{H81X} or a nearby proximal V_H (Supplementary Table 6). The skewed $V_H(D)J_H^+/V_H(D)J_H^-$ ratio in IGCR1^{+/-} B cells can be explained by frequent early formation of $V_{H81X}DJ_H$ rearrangements on the mutant allele. Thus, $V_{H81X}DJ_H^+$ rearrangements would exclude rearrangement of the wild-type allele but would be lost developmentally; leading to most peripheral B cells deriving from progenitors that formed productive $V_H(D)J_H$ rearrangements on the wild-type allele subsequent to $V_{H81X}(D)J_H^-$ rearrangements on the mutant allele (Supplementary Fig. 11d).

The extremely high representation of proximal V_H segments (for example, V_{H81X}) rearranged on the IGCR1-deleted allele might mask allelic inclusion because productive V_{H81X} rearrangements are selected against cellularly^{10,12,13}. Therefore, to examine further potential effects of IGCR1-deletion on allelic exclusion, we assayed the $V_H(D)J_H^+/DJ_H^-$ versus $V_H(D)J_H^+/V_H(D)J_H^-$ ratio of IGCR1^{-/-} hybridomas. Because both *Igh* alleles would be similarly biased for proximal V_H rearrangements in IGCR1^{-/-} B cells, one still would expect the 60/40 ratio if V_H -to- DJ_H recombination was feedback regulated (Supplementary Fig. 11e). However, we found an inverted ratio of 20/80 in IGCR1^{-/-} hybridomas (Fig. 4a), strongly suggesting that IGCR1-deleted alleles escape feedback regulation, at least for proximal V_H segments (Supplementary Fig. 11e). Because of the ambiguities of cellular selection against V_{H81X} and the lack of allotypically marked IGCR1-deleted alleles, we tested for escape from feedback inhibition by assaying for endogenous rearrangements in peripheral B cells from mice with a productive $V_H(D)J_H$ knock-in *Igh* allele (VB1-8 knock-in) that was IGCR1⁺ and a second allele that was IGCR1⁺ or IGCR1⁻. Notably, IGCR1^{+/-} VB1-8 knock-in B cells had a more than 20-fold increased level of V_{H7183} rearrangements compared to IGCR1^{+/+} VB1-8 knock-in B cells, but little if any change in the very low level rearrangement of distal V_H segments (Fig. 4b). Moreover, most rearrangements in IGCR1^{+/-} VB1-8 knock-in B cells were non-productive V_{H81X} rearrangements (Supplementary Fig. 11f), consistent with a lack of substantial allelic inclusion at the protein level in IGCR1^{+/-} F1 splenic B cells resulting from selection against V_{H81X} expression (Supplementary Figs 4c and 11a). We conclude that IGCR1 is required to allow feedback regulation of the most proximal V_H segments.

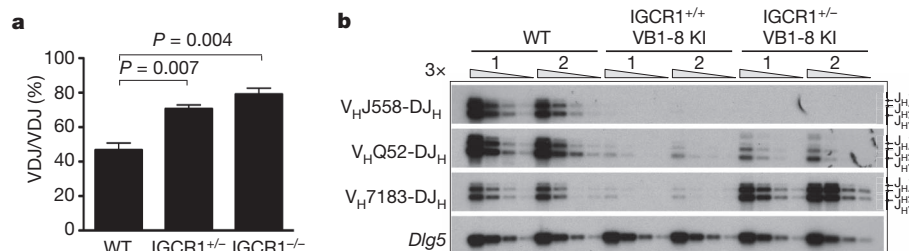


Figure 4 | IGCR1 is required to allow feedback regulation of proximal V_H -to- DJ_H recombination. **a**, Mean percentage of splenic B cells with $V_H(D)J_H$ rearrangements on both *Igh* alleles as determined by analyses of hybridomas from three independent sets of wild-type, IGCR1^{+/-} and IGCR1^{-/-} mice (Supplementary Table 5). Error bars represent standard deviation. P values

were calculated by Student's t -test. **b**, *Igh* $V_H(D)J_H$ rearrangements in splenic B cells from two independent wild-type and VB1-8 knock-in (KI) mice carrying either a wild-type (IGCR1^{+/+} VB1-8 KI) or an IGCR1-deleted (IGCR1^{-/-} VB1-8 KI) second allele. Bands corresponding to rearrangements to various J_H segments are indicated on right. *Dlg5* is the loading control.

IGCR1 mediates chromosomal *Igh* loops

We considered that IGCR1 might mediate *Igh* loops that would include iE μ and thereby modulate V(D)J recombination. The next CBEs downstream of IGCR1 are a set of 10, about 5 kb downstream of the 3' *Igh* regulatory region (3' *Igh* CBEs). To test for interactions between the IGCR1 and 3' *Igh* CBEs, we performed quantitative chromosome conformation capture (3C) assays on 129SV *Rag2*^{-/-}IGCR1^{+/+} and *Rag2*^{-/-}IGCR1^{-/-} A-MuLV-transformed pro-B lines. These analyses revealed interaction between the IGCR1 and 3' *Igh* CBE locales in *Rag2*^{-/-}IGCR1^{+/+} pro-B lines (Fig. 5a and Supplementary Fig. 12a), as found in another study³⁷. We also found this interaction in double-positive thymocytes (Supplementary Fig. 13). Notably, this interaction was eliminated in *Rag2*^{-/-}IGCR1^{-/-} pro-B lines (Fig. 5a and Supplementary Fig. 12a). We also found interactions between the iE μ locale and the IGCR1 and 3' *Igh* CBE locales in

Rag2^{-/-}IGCR1^{+/+} A-MuLV-transformed pro-B cells that were diminished in *Rag2*^{-/-}IGCR1^{-/-} pro-B lines (Fig. 5b and Supplementary Fig. 12b). Lastly, we found strong interactions between the iE μ and 3' *Igh* regulatory region, as reported for mature B cells³⁸, but these were not diminished by IGCR1 deletion (Fig. 5b). These studies demonstrate that IGCR1 mediates formation of 300-kb iE μ -containing *Igh* loops to the 3' *Igh* CBE locale in pro-B lines, with iE μ also being directly juxtaposed to the IGCR1 locale in an IGCR1-dependent manner, probably within the larger loop. As iE μ lacks CBEs, its interactions with the IGCR1 locale are probably mediated, at least in part, by factors other than CTCF.

Discussion

IGCR1, through its CBEs, mediates ordered and lineage-specific V_H-to-DJ_H recombination and balances proximal versus distal V_H

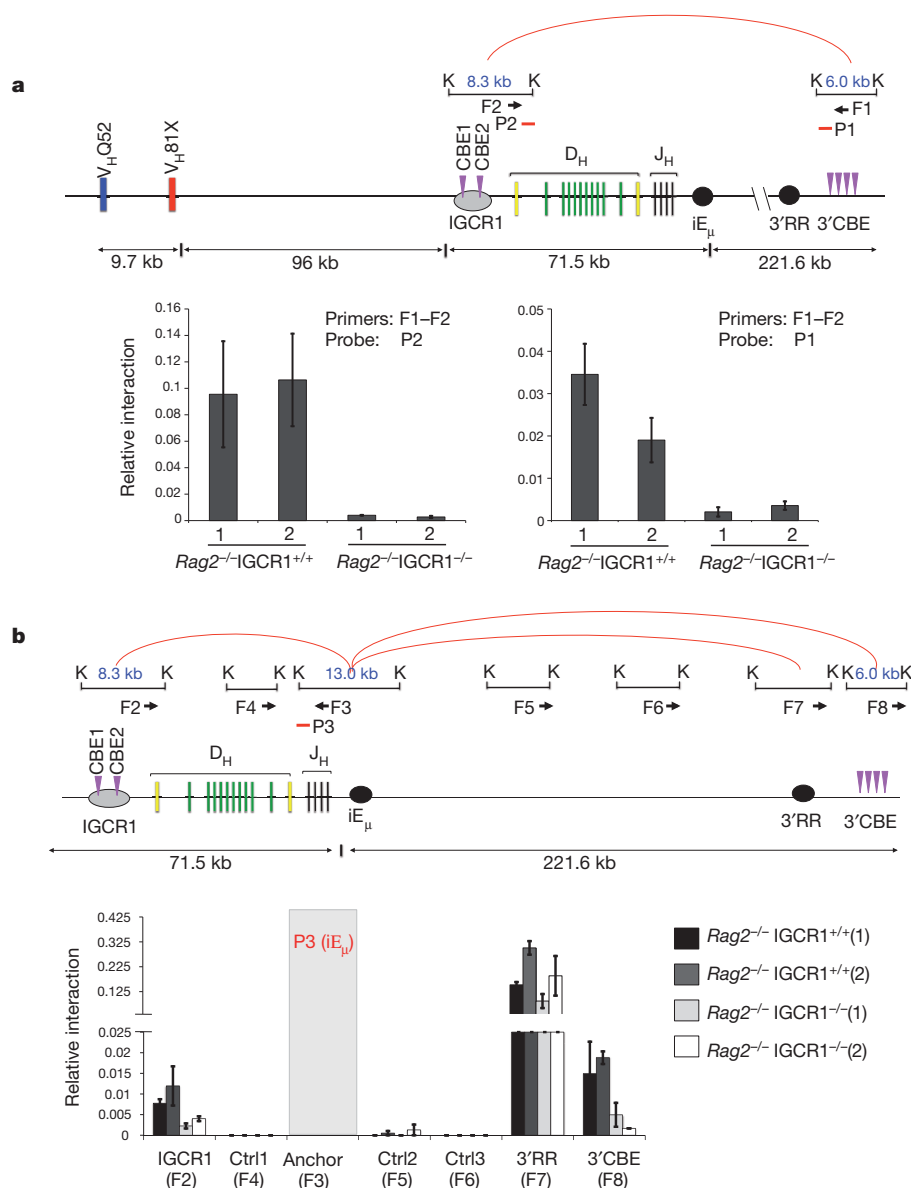


Figure 5 | IGCR1 mediates long-distance *Igh* chromosomal loops.

a, Schematic of chromosome interactions between IGCR1-containing and 3' *Igh* CBE-containing KpnI restriction fragments in 3C assays. Interactions between IGCR1 and 3' *Igh* CBE locales in 129SV *Rag2*^{-/-} and *Rag2*^{-/-}IGCR1^{-/-} A-MuLV-transformed pro-B cells were quantified by real-time PCR (Taqman) using probe P2 (left) and probe P1 (right). **b**, Schematic of chromosome interactions between iE μ -containing KpnI restriction fragment and indicated KpnI restriction fragments in other *Igh* locales. Interactions

between iE μ and IGCR1, iE μ and 3' *Igh* regulatory region (RR) locales, iE μ and 3' *Igh* CBE locales in *Rag2*^{-/-} and *Rag2*^{-/-}IGCR1^{-/-} A-MuLV-transformed pro-B cells were quantified by real-time PCR using a probe (P3) from the iE μ locale. F1–F8 indicate primers used for PCR. K indicates KpnI sites. Red arcs indicate interactions detected in *Rag2*^{-/-} cells. The average association frequency of three independent 3C experiments with two independent A-MuLV-transformed lines from each genotype is shown with standard deviation indicated.

rearrangement. Indeed, IGCR1 functions are required for an *Igh* allele to efficiently generate peripheral B cells. Notably, IGCR1 and its CBEs are not required for overall V_H -to-DJ_H recombination levels, but rather to decrease relative recombination of proximal V_H segments, particularly V_{H81X} . Inability of the dominant V_{H81X} to promote B-cell development probably leads to developmental defects associated with IGCR1 mutations. Yet, the enigmatic V_{H81X} is strongly conserved across mouse strains³⁹ and, correspondingly, has been suggested to have important roles in early antibody repertoires⁴⁰. Now, we find that IGCR1 has a key role in regulating V_{H81X} rearrangement. IGCR1 also is required to allow feedback regulation of proximal V_H -to-DJ_H rearrangements, implicating IGCR1 as a critical element for the allelic exclusion of V_{H81X} and other very proximal V_H segments. Our findings indicate that IGCR1 allows feedback by suppressing early, unordered proximal V_H rearrangement, providing the first evidence, to our knowledge, in support of a long-standing hypothesis that ordered V_H -to-DJ_H joining provides a means of mediating allelic exclusion^{5,6}. However, we found no evidence for loss of feedback regulation of distal V_H segments, in accordance with the proposal that locus contraction mediates their allelic exclusion²⁴.

Our findings show that IGCR1-mediated promotion of the utilization of V_H segments up to several megabases distant does not involve alterations in distal V_H transcription. In pro-B cells, *Igh* contraction promotes distal V_H usage^{23,24}. In the absence of certain transcription (for example, Pax-5 or YY1) or chromatin-modifying (for example, Ezh2) factors, distal V_H transcription is unimpaired but *Igh* contraction does not occur, diminishing distal V_H rearrangement. In such factor-deficient pro-B cells, transcription and rearrangement of proximal V_H segments does not increase^{8,41–43}, in contrast to the marked increases in IGCR1^{−/−} pro-B cells. This phenotypic difference is consistent with IGCR1 normalizing V_H repertoires via mechanisms other than *Igh* contraction. We suggest that IGCR1 promotes distal V_H usage indirectly by preventing premature proximal V_H rearrangement via insulating functions before contraction, thereby preserving DJ_H substrates for distal V_H rearrangement. The location of CBEs throughout the V_H portion of *Igh* led to the notion that recruitment of V_H segments into DJ_H recombination centres⁴⁴ subsequent to contraction is promoted via interaction of V_H and IGCR1 CBEs⁴⁵. Owing to the dominance of proximal V_H rearrangements on IGCR1-mutant alleles, assays for such putative IGCR1 functions require additional model systems.

IGCR1 CBEs suppress inappropriate transcription and rearrangement of proximal V_H segments 100 kb or more upstream. These suppressive functions are consistent with enhancer insulating functions of CBEs *in vitro*^{29,35}, which may relate to loop formation³⁵. We propose that IGCR1 CBEs mediate loops with downstream 3' *Igh* CBEs that segregate the D/J_H and V_H portions of *Igh* into separate regulatory domains during the D-to-J_H rearrangement stage of B-cell development, blocking activity of iEμ or other elements beyond IGCR1 (refs 17, 19; Supplementary Fig. 14a). Thus, inactivation of the IGCR1 CBEs allows transcriptional enhancing activity to extend to the proximal V_H segments promoting their premature rearrangement (Supplementary Fig. 14b). Notably, such activity does not appear to extend beyond the most proximal V_H segments, which may result from formation of new CBE-mediated loops to upstream V_H CBEs in the absence of IGCR1. In DJ_H-containing pro-B cells, IGCR1-insulating functions that prevent V_H -to-D_H rearrangements must be neutralized to allow V_H -to-DJ_H joining (Supplementary Fig. 14c). As CTCF binding to *Igh* CBEs does not vary with B-cell stage^{32,33}, other factors must modulate activity of bound CTCF within IGCR1 to allow for *Igh*-specific functions. Such factors might include CTCF modifications, interacting proteins such as cohesin^{32,33}, or CBE sequence context³⁵ and orientation^{46,47}. In addition, other putative binding elements within IGCR1 may recruit proteins, such as YY1, that have been implicated in modulating CTCF function³⁵.

METHODS SUMMARY

Mice. The targeting strategy and analysis of IGCR1-deleted and CBE-mutated embryonic stem (ES) cells is diagrammed in Supplementary Figs 2a and 3a (see Methods for details). The Institutional Animal Care and Use Committee of The Children's Hospital (Boston, Massachusetts) approved all animal work.

V(D)J rearrangement assays. PCR assays for D-to-J_H or V_H -to-DJ_H rearrangements were performed as described³⁰ (see Supplementary Table 1 for primers). Generation of B-cell hybridomas and V(D)J recombination analyses was performed as described⁴⁸.

RT-PCR and northern blot. RT-PCR and northern blotting assays for germline transcripts of *Igh* gene segments were performed as described³⁰ (primers for RT-PCR and northern blot probes are in Supplementary Table 1).

3C. 3C assays were performed as described⁴⁹.

ChIP-seq/ChIP-qPCR assays. Assays were done as described³².

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 November 2010; accepted 18 July 2011.

Published online 11 September 2011.

- Schatz, D. G. Antigen receptor genes and the evolution of a recombinase. *Semin. Immunol.* **16**, 245–256 (2004).
- Zhang, Y. *et al.* The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers. *Adv. Immunol.* **106**, 93–133 (2010).
- Yancopoulos, G. D. *et al.* Preferential utilization of the most J_H-proximal V_H gene segments in pre-B-cell lines. *Nature* **311**, 727–733 (1984).
- Perlot, T. & Alt, F. W. Cis-regulatory elements and epigenetic changes control genomic rearrangements of the IgH locus. *Adv. Immunol.* **99**, 1–32 (2008).
- Alt, F. W. *et al.* Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J.* **3**, 1209–1219 (1984).
- Jung, D., Giallourakis, C., Mostoslavsky, R. & Alt, F. W. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu. Rev. Immunol.* **24**, 541–570 (2006).
- Bates, J. G., Cado, D., Nolla, H. & Schissel, M. S. Chromosomal position of a V_H gene segment determines its activation and inactivation as a substrate for V(D)J recombination. *J. Exp. Med.* **204**, 3247–3256 (2007).
- Fuxa, M. *et al.* Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev.* **18**, 411–422 (2004).
- Reth, M. G., Ammirati, P., Jackson, S. & Alt, F. W. Regulated progression of a cultured pre-B-cell line to the B-cell stage. *Nature* **317**, 353–355 (1985).
- Melchers, F. *et al.* Repertoire selection by pre-B-cell receptors and B-cell receptors, and genetic control of B-cell development from immature to mature B cells. *Immunol. Rev.* **175**, 33–46 (2000).
- Malynn, B. A., Yancopoulos, G. D., Barth, J. E., Bona, C. A. & Alt, F. W. Biased expression of JH-proximal VH genes occurs in the newly generated repertoire of neonatal and adult mice. *J. Exp. Med.* **171**, 843–859 (1990).
- Decker, D. J., Boyle, N. E. & Klinman, N. R. Predominance of nonproductive rearrangements of VH81X gene segments evidences a dependence of B cell clonal maturation on the structure of nascent H chains. *J. Immunol.* **147**, 1406–1411 (1991).
- ten Boekel, E., Melchers, F. & Rolink, A. G. Changes in the V_H gene repertoire of developing precursor B lymphocytes in mouse bone marrow mediated by the pre-B cell receptor. *Immunity* **7**, 357–368 (1997).
- Yancopoulos, G. D. & Alt, F. W. Developmentally controlled and tissue-specific expression of unrearranged V_H gene segments. *Cell* **40**, 271–281 (1985).
- Yancopoulos, G. D., Blackwell, T. K., Suh, H., Hood, L. & Alt, F. W. Introduced T cell receptor variable region gene segments recombine in pre-B cells: evidence that B and T cells use a common recombinase. *Cell* **44**, 251–259 (1986).
- Corcoran, A. E. The epigenetic role of non-coding RNA transcription and nuclear organization in immunoglobulin repertoire generation. *Semin. Immunol.* **22**, 353–361 (2010).
- Feeney, A. Epigenetic regulation of V(D)J recombination. *Semin. Immunol.* **22**, 311–312 (2010).
- Subrahmanyam, R. & Sen, R. RAGs' eye view of the immunoglobulin heavy chain gene locus. *Semin. Immunol.* **22**, 337–345 (2010).
- Jhunjunwala, S., van Zelm, M. C., Peak, M. M. & Murre, C. Chromatin architecture and the generation of antigen receptor diversity. *Cell* **138**, 435–448 (2009).
- Abaratgui, I. & Krangel, M. S. Germline transcription: a key regulator of accessibility and recombination. *Adv. Exp. Med. Biol.* **650**, 93–102 (2009).
- Bergman, Y. & Cedar, H. Epigenetic control of recombination in the immune system. *Semin. Immunol.* **22**, 323–329 (2010).
- Sayegh, C. E., Jhunjunwala, S., Riblet, R. & Murre, C. Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes Dev.* **19**, 322–327 (2005).
- Jhunjunwala, S. *et al.* The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell* **133**, 265–279 (2008).
- Roldán, E. *et al.* Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nature Immunol.* **6**, 31–41 (2004).
- Pinaud, E. *et al.* The IgH locus 3' regulatory region: pulling the strings from behind. *Adv. Immunol.* **110**, 27–70 (2011).

26. Sakai, E., Bottaro, A., Davidson, L., Sleckman, B. P. & Alt, F. W. Recombination and transcription of the endogenous Ig heavy chain locus is effected by the Ig heavy chain intronic enhancer core region in the absence of the matrix attachment regions. *Proc. Natl Acad. Sci. USA* **96**, 1526–1531 (1999).
27. Perlot, T., Alt, F. W., Bassing, C. H., Suh, H. & Pinaud, E. Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc. Natl Acad. Sci. USA* **102**, 14362–14367 (2005).
28. Afshar, R., Pierce, S., Bolland, D. J., Corcoran, A. & Oltz, E. M. Regulation of IgH gene assembly: role of the intronic enhancer and 5'DQ52 region in targeting DHJH recombination. *J. Immunol.* **176**, 2439–2447 (2006).
29. Featherstone, K., Wood, A. L., Bowen, A. J. & Corcoran, A. E. The mouse immunoglobulin heavy chain V-D intergenic sequence contains insulators that may regulate ordered V(D)J recombination. *J. Biol. Chem.* **285**, 9327–9338 (2010).
30. Giallourakis, C. C. *et al.* Elements between the IgH variable (V) and diversity (D) clusters influence antisense transcription and lineage-specific V(D)J recombination. *Proc. Natl Acad. Sci. USA* **107**, 22207–22212 (2010).
31. Lin, Y. C. *et al.* A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nature Immunol.* **11**, 635–643 (2010).
32. Ebert, A. *et al.* The distal V_H gene cluster of the *Igh* locus contains distinct regulatory elements with pax5 transcription factor-dependent activity in pro-B cells. *Immunity* **34**, 175–187 (2011).
33. Degner, S. C., Wong, T. P., Jankevicius, G. & Feeney, A. J. Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. *J. Immunol.* **182**, 44–48 (2009).
34. Williams, A. & Flavell, R. A. The role of CTCF in regulating nuclear organization. *J. Exp. Med.* **205**, 747–750 (2008).
35. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
36. Bulger, M. & Groudine, M. Enhancers: the abundance and function of regulatory sequences beyond promoters. *Dev. Biol.* **339**, 250–257 (2010).
37. Degner, S. C. *et al.* CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the *Igh* locus and antisense transcription in pro-B cells. *Proc. Natl Acad. Sci. USA* **108**, 9566–9571 (2011).
38. Wuerffel, R. *et al.* S-S synapsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. *Immunity* **27**, 711–722 (2007).
39. Hirano, S. L. *et al.* Identity of IGHV-7183.1 (V81x) coding and recombination signal sequences among wild-derived mice. *Immunogenetics* **53**, 54–58 (2001).
40. Hardy, R. R. & Hayakawa, K. B cell development pathways. *Annu. Rev. Immunol.* **19**, 595–621 (2001).
41. Hesslein, D. G. *et al.* Pax5 is required for recombination of transcribed, acetylated, 5' IgH V gene segments. *Genes Dev.* **17**, 37–42 (2003).
42. Su, I. H. *et al.* Ezh2 controls B cell development through histone H3 methylation and *Igh* rearrangement. *Nature Immunol.* **4**, 124–131 (2003).
43. Liu, H. *et al.* Yin Yang 1 is a critical regulator of B-cell development. *Genes Dev.* **21**, 1179–1189 (2007).
44. Ji, Y. *et al.* The *in vivo* pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell* **141**, 419–431 (2010).
45. Degner-Leisso, S. C. & Feeney, A. J. Epigenetic and 3-dimensional regulation of V(D)J rearrangement of immunoglobulin genes. *Semin. Immunol.* **22**, 346–352 (2010).
46. MacPherson, M. J. & Sadowski, P. D. The CTCF insulator protein forms an unusual DNA structure. *BMC Mol. Biol.* **11**, 101 (2010).
47. Kyrchanova, O., Chetverina, D., Maksimenko, O., Kuliyev, A. & Georgiev, P. Orientation-dependent interaction between *Drosophila* insulators is a property of this class of regulatory elements. *Nucleic Acids Res.* **36**, 7019–7028 (2008).
48. Dudley, D. D. *et al.* Impaired V(D)J recombination and lymphocyte development in core RAG1-expressing mice. *J. Exp. Med.* **198**, 1439–1450 (2003).
49. Hagege, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nature Protocols* **2**, 1722–1733 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Fujiwara and P.-Y. Huang for generating chimaeric mice. This work was supported by NIH grants RO1 AI20047 (to F.W.A.), RO1 HL48702 and AI40227 (to M.S.S.), CA054198-20 (to C.M.) and K08 AI070839 (to C.C.G.). M.B. was supported by the Austrian GEN-AU initiative and Boehringer Ingelheim. C.G. is supported by an Irvington Institute Postdoctoral Fellowship from the Cancer Research Institute. C.V. is supported by a Marie Curie Fellowship. C.B. is supported by an EMBO fellowship. F.W.A. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions C.G., H.S.Y., A.F., C.C.G. and F.W.A. conceived, designed and/or performed most experiments, interpreted most results, and wrote the manuscript. With respect to other authors, S.J. performed experiments for Figs 2, 3, 4 and Supplementary Fig. 11 and contributed to interpretation of the data; A.E. and M.B. contributed the work in Fig. 2c and Supplementary 7a–c. C.V., J.G.B. and M.S.S. contributed the work in Supplementary Fig. 10; and C.B. and C.M. performed FISH experiments on IGCR1^{-/-} cells that helped frame aspects of the discussion and models; all of these authors contributed to polishing the manuscript. All other authors provided technical assistance with various experiments or data analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.W.A. (alt@enders.tch.harvard.edu) or C.C.G. (cgiallourakis@partners.org).

METHODS

Generation of IGCRI-deleted mice. A targeting construct was designed to replace IGCRI (4.1 kb) with a NeoR gene cassette oriented in the direction from the D clusters to V clusters (Supplementary Fig. 2a). A 4.3-kb arm upstream of IGCRI and a 2.9-kb arm downstream of IGCRI were PCR amplified (see Supplementary Table 1 for primers) from TC1 embryonic stem (ES) cell DNA (129 strain) and cloned into the pLNTK targeting vector in the desired orientation. The targeting construct was then electroporated into TC1 ES cells, and successful targeting assessed by Southern blot analyses using StuI- or SpeI-digested genomic DNA and upstream or downstream genomic probes as outlined in detail in Supplementary Fig. 2a. Three independently targeted ES clones were subjected to adenovirus-mediated Cre deletion to remove the NeoR gene and injected for *Rag2*-deficient blastocyst complementation (RDBC)⁵⁰ or for germline transmission.

Generation of IGCRI CBE-mutated mice. Two 4.2-kb DNA fragments consecutively located over the IGCRI region were PCR amplified (see Supplementary Table 1 for primers) and cloned into a pGEM-T easy (Promega) vector (Supplementary Fig. 3a). One fragment included CBE1 and the other included CBE2. PCR site-directed mutagenesis was used to introduce scrambled mutations of the 20-bp CBE1 and 19-bp CBE2 sites in each arm, respectively (see Supplementary Table 1 for primers). Restriction endonuclease recognition sites were incorporated into the mutated CBE sequences (NheI for upstream and SpeI for downstream arms). Then, these two DNA fragments were cloned into a targeting vector pLNTK as upstream and downstream arms. The targeting construct was electroporated into TC1 ES cells, and successfully targeted clones, including no mutations (*loxP* insertion control) and the CBE1 and 2 double mutation were assessed by Southern blot analyses using StuI-, SpeI- or HindIII/NheI-digested genomic DNA with appropriate probes (Supplementary Fig. 3b). Two independently targeted clones were subjected to adenovirus-mediated Cre deletion to remove the NeoR gene and injected for RDBC or for germline transmission. For RDBC, sorted double-positive T cells from chimaeras were genotyped by PCR and restriction enzyme digestion (see Supplementary Fig. 3c). Wild-type 129SV and C57BL/6 mice were purchased from Jackson Laboratory. *Rag2*-deficient mice on a 129 background were purchased from Taconic.

Electrophoretic mobility shift assay. Probes were prepared by annealing complementary oligonucleotides (Supplementary Table 1). Annealed oligonucleotides were purified on 4% agarose gels and end-labelled with ³²P-γ-ATP. Nuclear extracts were prepared from *Rag2*-deficient A-MuLV-transformed pro-B cell lines. Electrophoretic mobility shift assay (EMSA) reactions were conducted in a mixture of 5% glycerol, 150 mM KCl, 20 mM HEPES, pH 7.9, 5 mM MgCl₂, 1 mM dithiothreitol (DTT), 0.5% Triton X-100, 400 ng poly(dG-dC). 2 μg of nuclear extract was incubated with anti-CTCF or anti-IgG antibodies at 4 °C for 20 min and labelled probes and/or competitor un-labelled probes were added to the reactions. The reactions were electrophoresed with 0.5× TBE buffer (89 mM Tris Base, 89 mM boric acid, 2 mM EDTA, pH 8.0) at 30 V, and visualized by autoradiography.

V(D)J recombination assays. Genomic DNA was purified from sorted bone marrow pro-B (IgM⁺B220⁺CD43⁺), pre-B (IgM⁺B220⁺CD43⁺) cells, and splenic mature B (IgM⁺B220⁺CD43⁺) cells, and double-positive T (B220⁺CD4⁺CD8⁺) cells. Fivefold serial dilutions of genomic DNA (200 ng, 40 ng, 8 ng) was used to perform PCR to analyse V(D)J rearrangements. Primers used in this assay are listed in Supplementary Table 1. Primers flanking exon 6 of the *Dlg5* gene were used as a loading control under the same conditions. Vκ-to-Jκ rearrangement PCRs were performed to confirm specificity of sorted B-cell populations, and to exclude potential B-cell contamination during double-positive T-cell analysis. PCR products were gel electrophoresed and transferred to determine V(D)J recombination by Southern blotting using radiolabelled oligonucleotide probes (see Supplementary Table 1 for sequences) and visualized by autoradiography.

RNA isolation and RT-PCR. Total RNA was isolated using Trizol (Invitrogen). One microgram of RNA was used to generate cDNA with reverse transcriptase Superscript III (Invitrogen) with random hexamers according to manufacturer's protocols. Approximately, 1/40 of the reverse-transcription-generated cDNA was analysed by PCR. Primers that were used for PCR are provided in Supplementary Table 1.

Intracytoplasmic staining. Intracytoplasmic staining was performed as described previously⁵¹. Briefly, splenic B cells from F1 mice with a wild-type *Igmb* allele and an IGCRI-deleted *Igmb* allele were purified by MACS paramagnetic beads

following the manufacturer's protocol and stimulated for 4 days with LPS. Cells were fixed, permeabilized and then stained with FITC-labelled anti-mouse IgM^a and biotin-labelled anti-mouse IgM^b revealed by streptavidin-conjugated Texas Red. Cells were examined using a fluorescent microscope for IgM^a and IgM^b allotypic expressers.

Hybridoma assay and Southern blot. Splenic B cells were isolated from wild-type, IGCRI^{+/-} and IGCRI^{-/-} mice, and fused with NS1 cells after stimulation with 25 ng ml⁻¹ IL-4 and 500 ng ml⁻¹ anti-CD40 antibody for 4 days in culture. Hybridoma cells were plated and selected in HAT medium as previously described⁴⁸. Genomic DNA from hybridomas was isolated and digested with StuI to determine V(D)J rearrangement configurations by Southern blot (Supplementary Fig. 11). DNA from the clones that showed V_H(D)_H rearrangement on both alleles was subjected to PCR using an upstream V_H primer (specific to V_{HJ558}, V_{HQ52} or V_{H7183} V_H gene families) and a downstream J_{H4} primer, and the amplified junctions were cloned and sequenced to identify productive and non-productive V_H(D)_H rearrangements.

Transgenic mice. IGCRI^{+/-} mice were bred with the mutant mice harbouring a pre-assembled *Igh* VB1-8DJ_{H4} allele⁵². B cells were purified by MACS paramagnetic beads from VB1-8DJ_{H4} knock-in mice with either wild-type IGCRI (IGCRI^{+/+} VB1-8 knock-in) or IGCRI deletion (IGCRI^{+/-} VB1-8 knock-in) on the other *Igh* allele. Genomic DNA was isolated and V(D)J rearrangement of V_{H7183}, V_{HQ52}, and V_{HJ558} segments were amplified as described in Supplementary Fig. 6a.

ChIP. Rabbit polyclonal antibodies recognizing the following histone tail modifications were used: H3K9ac (Millipore; 07-352), H3K4me2 (Millipore; 07-030) and H3K4me3 (Diagenode; pAB-003-050). ChIP analysis and ChIP sequencing of A-MuLV-transformed pro-B cells was performed as described³². The sequence reads obtained by paired-end Solexa sequencing with a read length of 76 nucleotides were mapped to the 129SV mouse reference genome. The ChIP-qPCR analysis was performed by quantifying the precipitated DNA on a MyiQ instrument (Bio-Rad) as described³². The amount of precipitated DNA was determined as percentage relative to input DNA to obtain relative enrichment compared to the precipitated DNA of the control *Bcar3* enhancer⁵³. Tenfold dilutions of input material were used to generate a standard curve, and ChIP samples were quantified relative to input using the iQ5 software. The oligonucleotides used for real-time PCR analysis are shown in Supplementary Table 1.

3C. The 3C assays were performed essentially as previously described⁴⁹. Briefly, 2 × 10⁷ cells were cross-linked with 1% formaldehyde for 10 min. The reaction was quenched with glycine (0.125 M). Cells were lysed in 10 mM Tris pH.8, 10 mM NaCl and 0.2% NP-40 followed by 15 strokes using a dounce homogenizer. The resulting nuclei were washed in restriction enzyme buffer, resuspended with the same buffer containing 0.3% SDS, and incubated for 1 h at 37 °C. To sequester SDS, 2% Triton X-100 was added, and incubated for 1 h at 37 °C. 400 U KpnI was added and incubated overnight at 37 °C. KpnI was inactivated with 1.6% SDS and incubated for 25 min at 68 °C. The samples were ligated in ligation buffer (50 mM Tris, 10 mM MgCl₂, 1% Triton X-100, 100 mM DTT and 0.1 M ATP) with T4 DNA ligase overnight at 16 °C. The crosslinks within 3C library products were reversed and the DNA purified by overnight treatment with proteinase K at 65 °C as per assay protocol. Quantitative real-time PCR using a standard curve was conducted to measure the frequency of the 3C products within each sample. Standard curves for 3C assays were generated using BACs containing the IGCRI, iEμ and 3' *Igh* CBE locales within the *Igh* locus (RP23-38K22, RP23-334P5 and RP24-275024) that were KpnI-digested and then religated to generate all possible 3C products within the locus. Taqman applied real-time PCR was used to determine a 3C frequency by averaging the amount of 3C products produced for a given amplicon and dividing that value by the amount of loading control determined by loading control amplicon (see Supplementary Table 1).

50. Chen, J. *et al.* Mutations of the intronic IgH enhancer and its flanking sequences differentially affect accessibility of the JH locus. *EMBO J.* **12**, 4635–4645 (1993).
51. Barreto, V. & Cumano, A. Frequency and characterization of phenotypic Ig heavy chain allelically included IgM-expressing B cells in mice. *J. Immunol.* **164**, 893–899 (2000).
52. Sonoda, E. *et al.* B cell development under the condition of allelic inclusion. *Immunology* **6**, 225–233 (1997).
53. McManus, S. *et al.* The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. *EMBO J.* **30**, 2388–2404 (2011).

Frequent pathway mutations of splicing machinery in myelodysplasia

Kenichi Yoshida^{1*}, Masashi Sanada^{1*}, Yuichi Shiraishi^{2*}, Daniel Nowak^{3*}, Yasunobu Nagata^{1*}, Ryo Yamamoto⁴, Yusuke Sato¹, Aiko Sato-Otsubo¹, Ayana Kon¹, Masao Nagasaki⁵, George Chalkidis⁶, Yutaka Suzuki⁷, Masashi Shiosaka¹, Ryoichiro Kawahata¹, Tomoyuki Yamaguchi⁸, Makoto Otsu⁴, Naoshi Obara⁹, Mamiko Sakata-Yanagimoto⁹, Ken Ishiyama¹⁰, Hiraku Mori¹¹, Florian Nolte³, Wolf-Karsten Hofmann³, Shuichi Miyawaki¹⁰, Sumio Sugano⁷, Claudia Haerlach¹², H. Phillip Koeffler^{13,14}, Lee-Yung Shih¹⁵, Torsten Haerlach¹², Shigeru Chiba⁹, Hiromitsu Nakauchi^{4,8}, Satoru Miyano^{2,6} & Seishi Ogawa¹

Myelodysplastic syndromes and related disorders (myelodysplasia) are a heterogeneous group of myeloid neoplasms showing deregulated blood cell production with evidence of myeloid dysplasia and a predisposition to acute myeloid leukaemia, whose pathogenesis is only incompletely understood. Here we report whole-exome sequencing of 29 myelodysplasia specimens, which unexpectedly revealed novel pathway mutations involving multiple components of the RNA splicing machinery, including *U2AF35*, *ZRSR2*, *SRSF2* and *SF3B1*. In a large series analysis, these splicing pathway mutations were frequent (~45 to ~85%) in, and highly specific to, myeloid neoplasms showing features of myelodysplasia. Conspicuously, most of the mutations, which occurred in a mutually exclusive manner, affected genes involved in the 3'-splice site recognition during pre-mRNA processing, inducing abnormal RNA splicing and compromised haematopoiesis. Our results provide the first evidence indicating that genetic alterations of the major splicing components could be involved in human pathogenesis, also implicating a novel therapeutic possibility for myelodysplasia.

Myelodysplastic syndromes (MDS) and related disorders (myelodysplasia) comprise a group of myeloid neoplasms characterized by deregulated, dysplastic blood cell production and a predisposition to acute myeloid leukaemia (AML)¹. Although the prevalence of MDS has not been determined precisely, more than 10,000 people are estimated to develop myelodysplasia annually in the United States². Their indolent clinical course before leukaemic transformation and ineffective haematopoiesis with evidence of myeloid dysplasia indicate a pathogenesis distinct from that involved in *de novo* AML. Currently, a number of gene mutations and cytogenetic changes have been implicated in the pathogenesis of MDS, including mutations of *RAS*, *TP53* and *RUNX1*, and more recently *ASXL1*, *c-CBL*, *DNMT3A*, *IDH1/2*, *TET2* and *EZH2* (ref. 3). Nevertheless, mutations of this set of genes do not fully explain the pathogenesis of MDS because they are also commonly found in other myeloid malignancies and roughly 20% of MDS cases have no known genetic changes (ref. 4 and unpublished data). In particular, the genetic alterations responsible for the dysplastic phenotypes and ineffective haematopoiesis of myelodysplasia are poorly understood. Meanwhile, the recent development of massively parallel sequencing technologies has provided an expanded opportunity to discover genetic changes across the entire genomes or protein-coding sequences in human cancers at a single-nucleotide level^{5–10}, which could be successfully applied to the genetic analysis of myelodysplasia to obtain a better understanding of its pathogenesis.

Overview of genetic alterations

In this study, we performed whole-exome sequencing of paired tumour/control DNA from 29 patients with myelodysplasia (Supplementary Table 1). Although incapable of detecting non-coding mutations and gene rearrangements, the whole-exome approach is a well-established strategy for obtaining comprehensive registries of protein-coding mutations at low cost and high performance. With a mean coverage of 133.8, 80.4% of the target sequences were analysed at more than $\times 20$ depth on average (Supplementary Fig. 1). All the candidates for somatic mutations ($N = 497$) generated through our data analysis pipeline were subjected to validation using Sanger sequencing (Supplementary Methods I and Supplementary Fig. 2). Finally, 268 non-synonymous somatic mutations were confirmed with an overall true positive rate of 53.9% (Supplementary Fig. 3), including 206 missense, 25 nonsense, and 10 splice site mutations, and 27 frameshift-causing insertions/deletions (indels) (Supplementary Fig. 4). The mutation rate of 9.2 (0–21) per sample was significantly lower than that in solid tumours (16.2–302)^{7,11,12} and multiple myeloma (32.4)⁶, but was comparable to that in AML (7.3–13)^{13–15} and chronic lymphocytic leukaemia (11.5)¹⁶. Combined with the genomic copy number profile obtained by single nucleotide polymorphism (SNP) array karyotyping, this array of somatic mutations provided a landscape of myelodysplasia genomes (Supplementary Fig. 5)^{17,18}.

¹Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. ²Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ³Department of Hematology and Oncology, Medical Faculty Mannheim of the University of Heidelberg, 1-3 Theodor-Kutzer-Ufer, Mannheim 68167, Germany. ⁴Division of Stem Cell Therapy, Center for Stem Cell Biology and Regenerative Medicine, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ⁵Laboratory of Functional Genomics, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ⁶Laboratory of Sequence Data Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ⁷Division of Systems Biomedical Technology, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ⁸Nakauchi Stem Cell and Organ Regeneration Project, Exploratory Research for Advanced Technology, Japan Science and Technology Agency, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ⁹Department of Hematology, Institute of Clinical Medicine, University of Tsukuba, 1-1-1 Tennodai, Tsukuba-shi, Ibaraki, 305-8571, Japan. ¹⁰Division of Hematology, Tokyo Metropolitan Ohtsuka Hospital, 2-8-1 Minami-Ohtsuka, Toshima-ku, Tokyo 170-0005, Japan. ¹¹Division of Hematology, Internal Medicine, Showa University Fujigaoka Hospital, 1-30 Fujigaoka, Aoba-ku, Yokohama, Kanagawa 227-8501, Japan. ¹²Munich Leukemia Laboratory, Max-Lebsche-Platz 31, Munich 81377, Germany. ¹³Hematology/Oncology, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, California 90048, USA. ¹⁴National University of Singapore, Cancer Science Institute of Singapore, 28 Medical Drive, Singapore 117456, Singapore. ¹⁵Division of Hematology-Oncology, Department of Internal Medicine, Chang Gung Memorial Hospital, Chang Gung University, 199 Tung Hwa North Rd, Taipei 105, Taiwan.

*These authors contributed equally to this work.

Novel gene targets in myelodysplasia

The list of the somatic mutations (Supplementary Table 2) included most of the known gene targets in myelodysplasia with similar mutation frequencies to those previously reported, indicating an acceptable sensitivity of the current study. The mutations of the known gene targets, however, accounted for only 12.3% of all detected mutations ($N = 33$), and the remaining 235 mutations involved previously unreported genes. Among these, recurrently mutated genes in multiple cases are candidate targets of particular interest, for which high mutation rates are expected in general populations. In fact, 8 of the 12 recurrently mutated genes were among the well-described gene targets in myelodysplasia (Supplementary Table 3). However, what immediately drew our attention were the recurrent mutations involving *U2AF35* (also known as *U2AF1*), *ZRSR2* and *SRSF2* (SC35), because they belong to the common pathway known as RNA splicing. Including an additional three genes mutated in single cases (*SF3A1*, *SF3B1* and *PRPF40B*), six components of the splicing machinery were mutated in 16 out of the 29 cases (55.2%) in a mutually exclusive manner (Fig. 1, Supplementary Fig. 6 and Supplementary Table 2).

Frequent mutations in splicing machinery

RNA splicing is accomplished by a well-ordered recruitment, rearrangement and/or disengagement of a set of small nuclear ribonucleoprotein (snRNP) complexes (U1, U2, and either U4/5/6 or U11/12), as well as many other protein components onto the pre-mRNAs. Notably, the mutated components of the spliceosome were all engaged in the initial steps of RNA splicing, except for *PRPF40B*, whose functions in RNA splicing are poorly defined. Making physical interactions with SF1 and a serine/arginine-rich (SR) protein, such as *SRSF1* or *SRSF2*, the U2 auxiliary factor (U2AF) that consists of the *U2AF65* (*U2AF2*)–*U2AF35* heterodimer, is involved in the recognition of the 3' splice site (3'SS) and its nearby polypyrimidine tract, which is thought to be required for the subsequent recruitment of the U2 snRNP, containing *SF3A1* as well as *SF3B1*, to establish the splicing A complex (Fig. 1)¹⁹. *ZRSR2* (or *Urp*), is another essential component of the splicing machinery. Showing a close structural similarity to *U2AF35*, *ZRSR2* physically interacts with *U2AF65*, as well as *SRSF1* and *SRSF2*, with a distinct function from its homologue, *U2AF35* (ref. 20).

To confirm and extend the initial findings in the whole-exome sequencing, we studied mutations of the above six genes together with

three additional spliceosome-related genes, including *U2AF65*, *SF1* and *SRSF1*, in a large series of myeloid neoplasms ($N = 582$) using a high-throughput mutation screen of pooled DNA followed by confirmation/identification of candidate mutations (refs 21 and 22 and Supplementary Methods II).

In total, 219 mutations were identified in 209 out of the 582 specimens of myeloid neoplasms through validating 313 provisional positive events in the pooled DNA screen (Supplementary Tables 4 and 5). The mutations among four genes, *U2AF35* ($N = 37$), *SRSF2* ($N = 56$), *ZRSR2* ($N = 23$) and *SF3B1* ($N = 79$), explained most of the mutations with much lower mutational rates for *SF3A1* ($N = 8$), *PRPF40B* ($N = 7$), *U2AF65* ($N = 4$) and *SF1* ($N = 5$) (Fig. 2). Mutations of the splicing machinery were highly specific to diseases showing myelodysplastic features, including MDS either with (84.9%) or without (43.9%) increased ring sideroblasts, chronic myelomonocytic leukaemia (CMML) (54.5%), and therapy-related AML or AML with myelodysplasia-related changes (25.8%), but were rare in *de novo* AML (6.6%) and myeloproliferative neoplasms (MPN) (9.4%) (Fig. 3a). The mutually exclusive pattern of the mutations in these splicing pathway genes was confirmed in this large case series, suggesting a common impact of these mutations on RNA splicing and the pathogenesis of myelodysplasia (Fig. 3b). The frequencies of mutations showed significant differences across disease types. Surprisingly, *SF3B1* mutations were found in the majority of the cases with MDS characterized by increased ring sideroblasts, that is, refractory anaemia with ring sideroblasts (RARS) (19/23 or 82.6%) and refractory cytopenia with multilineage dysplasia with $\geq 15\%$ ring sideroblasts (RCMD-RS) (38/50 or 76%) with much lower mutation frequencies in other myeloid neoplasms. RARS and RCMD-RS account

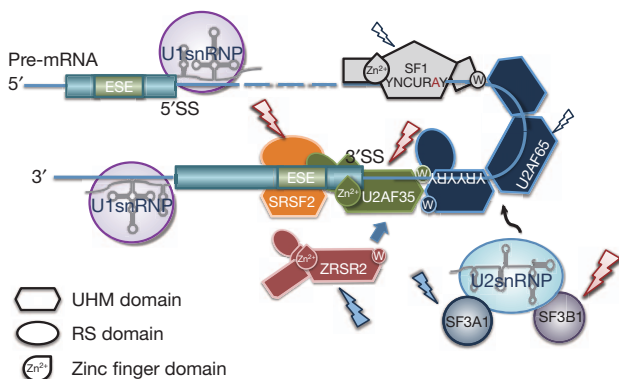


Figure 1 | Components of the splicing E/A complex mutated in myelodysplasia. RNA splicing is initiated by the recruitment of U1 snRNP to the 5'SS. SF1 and the larger subunit of the U2 auxiliary factor (U2AF), *U2AF65*, bind the branch point sequence (BPS) and its downstream polypyrimidine tract, respectively. The smaller subunit of U2AF (*U2AF35*) binds to the AG dinucleotide of the 3'SS, interacting with both *U2AF65* and a SR protein, such as *SRSF2*, through its UHM and RS domain, comprising the earliest splicing complex (E complex). *ZRSR2* also interacts with U2AF and SR proteins to perform essential functions in RNA splicing. After the recognition of the 3'SS, U2 snRNP, together with *SF3A1* and *SF3B1*, is recruited to the 3'SS to generate the splicing complex A. The mutated components in myelodysplasia are indicated by arrows.

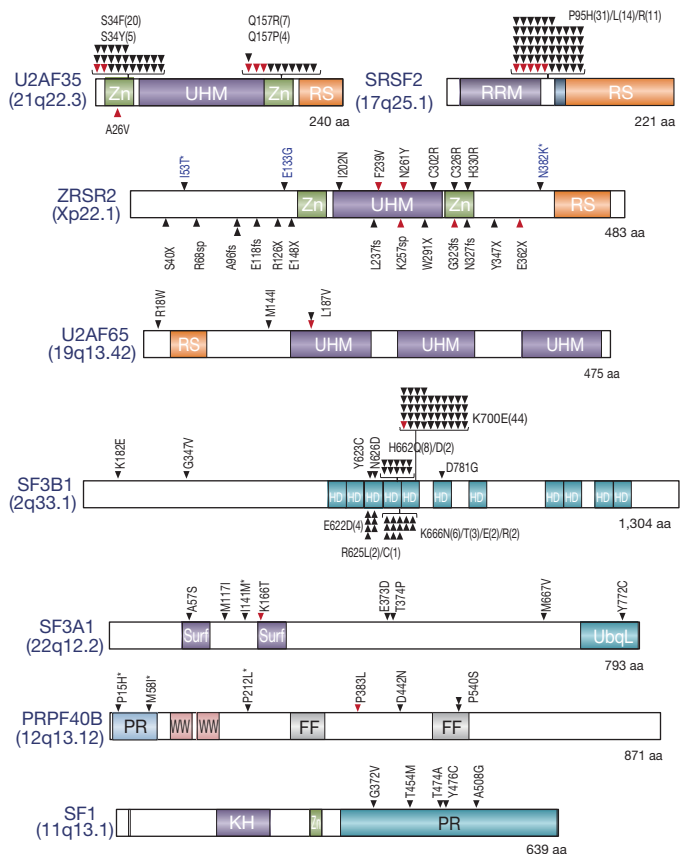


Figure 2 | Mutations of multiple components of the splicing machinery. Each mutation in the eight spliceosome components is shown with an arrowhead. Confirmed somatic mutations are discriminated by red arrows. Known domain structures are shown in coloured boxes as indicated. Mutations predicted as SNPs by MutationTaster (<http://www.mutationtaster.org/>) are indicated by asterisks. The number of each mutation is indicated in parenthesis. *ZRSR2* mutations in females are shown in blue.

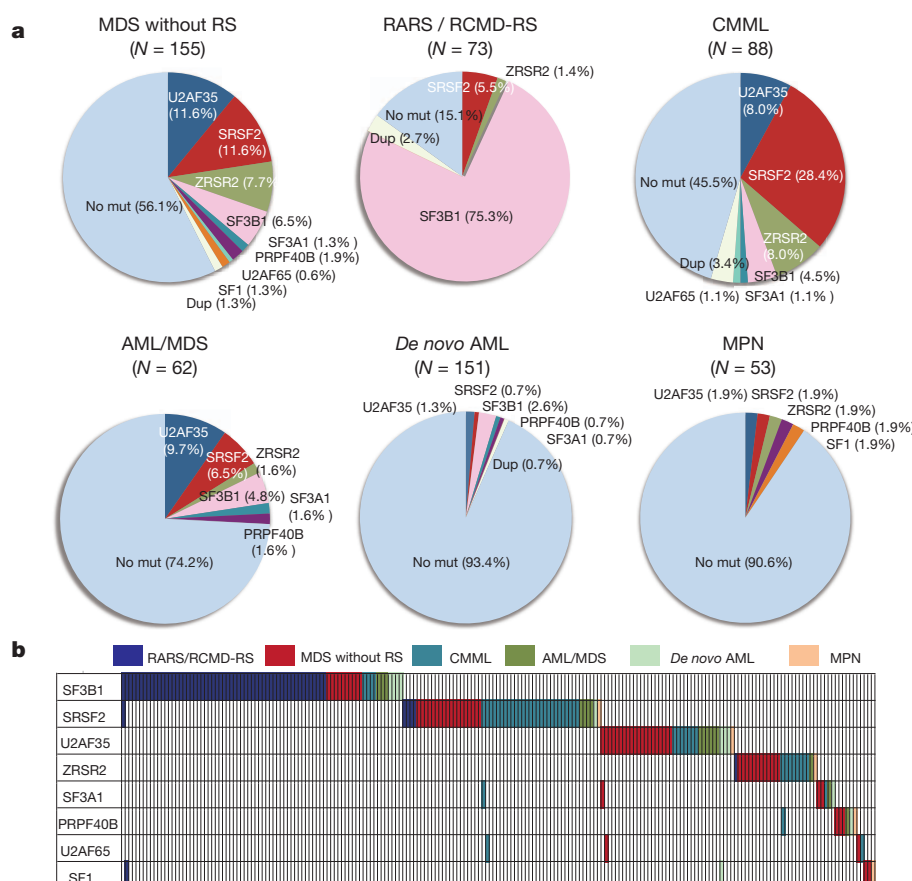


Figure 3 | Frequencies and distribution of spliceosome pathway gene mutations in myeloid neoplasms. a, Frequencies of spliceosome pathway mutations among 582 cases with various myeloid neoplasms. **b**, Distribution of mutations in eight spliceosome genes, where diagnosis of each sample is shown by indicated colours.

for 4.3% and 12.9% of MDS cases, respectively, where deregulated iron metabolism has been implicated in the development of refractory anaemia²³. With such high mutation frequencies and specificity, the *SF3B1* mutations were thought to be almost pathognomonic to these MDS subtypes characterized by increased ring sideroblasts, and strongly implicated in the pathogenesis of MDS in these categories. Less conspicuously but significantly, *SRSF2* mutations were more frequent in CMML cases (Fig. 3 and Supplementary Table 4). Thus, although commonly involving the E/A splicing complexes, different mutations may still have different impacts on cell functions, contributing to the determination of discrete disease phenotypes. For example, studies have demonstrated that *SRSF2* was also involved in the regulation of DNA stability and that depletion of *SRSF2* can lead to genomic instability²⁴. Of interest in this context, regardless of disease subtypes, samples with *SRSF2* mutations were shown to have significantly more mutations of other genes compared with *U2AF35* mutations ($P = 0.001$, multiple regression analysis) (Supplementary Table 6 and Supplementary Fig. 7).

Notably, with a rare exception of A26V in a single case, the mutations of *U2AF35* exclusively involved two highly conserved amino acid positions (S34 or Q157) within the amino- and the carboxyl-terminal zinc finger motifs flanking the U2AF homology motif (UHM) domain. *SRSF2* mutations exclusively occurred at P95 within an intervening sequence between the RNA recognition motif (RRM) and arginine/serine-rich (RS) domains (Fig. 2 and Supplementary Figs 8 and 9). Similarly, *SF3B1* mutations predominantly involved K700 and, to a lesser extent, K666, H662 and E622, which are also conserved across species (Fig. 2 and Supplementary Fig. 10). The involvement of recurrent amino acid positions in these spliceosome genes strongly indicated a gain-of-function nature of these mutations, which has been a well-documented scenario in other oncogenic mutations²⁵. On the other hand, the 23 mutations in *ZRSR2* (Xp22.1) were widely distributed along the entire coding region (Fig. 2). Among these, 14 mutations were nonsense or frameshift changes, or involved splicing donor/acceptor

sites that caused either a premature truncation or a large structural change of the protein, leading to loss-of-function. Combined with their strong male preference for the mutation (14/14 cases), *ZRSR2* most likely acts as a tumour suppressor gene with an X-linked recessive mode of genetic action. The remaining nine *ZRSR2* mutations were missense changes and found in both males (six cases) and females (three cases), whose somatic origin was only confirmed in two cases. However, neither the dbSNP database (build131 and 132) nor the 1000 Genomes database (May 2011 snp calls) contained these missense nucleotides, suggesting that many, if not all, of these missense changes are likely to represent functional somatic changes, especially those found in males. Interrogation of these hot spots for mutations in *U2AF35* and *SRSF2* found no mutations among lymphoid neoplasms, including acute lymphoblastic leukaemia ($N = 24$) or non-Hodgkin's lymphoma ($N = 87$) (data not shown).

RNA splicing and spliceosome mutations

Because the splicing mutations in myelodysplasia widely and specifically affect the major components of the splicing complexes E/A in a mutually exclusive manner, the common consequence of these mutations is logically the impaired recognition of 3'SSs that would lead to the production of aberrantly spliced mRNA species. To appreciate this and also to gain an insight into the biological/biochemical impact of these splicing mutations, we expressed the wild-type and the mutant (S34F) *U2AF35* in HeLa cells using retrovirus-mediated gene transfer with enhanced green fluorescent protein (EGFP) marking (Fig. 4a and Supplementary Methods III) and examined their effects on gene expression in these cells using GeneChip Human genome U133 plus 2.0 arrays (Affymetrix), followed by gene set enrichment analysis (GSEA) (Supplementary Methods IV)²⁶. Intriguingly, the GSEA disclosed a significant enrichment of the genes on the non-sense-mediated mRNA decay (NMD) pathway among the significantly upregulated genes in mutant *U2AF35*-transduced HeLa cells (Fig. 4b, Supplementary Fig. 11a and Supplementary Table 7), which was

confirmed by quantitative polymerase chain reactions (qPCR) (Fig. 4c and Supplementary Methods 5V). A similar result was also observed for the gene expression profile of an MDS-derived cell line (TF-1) transduced with the S34F mutant (Supplementary Figs 11b, c). The NMD activation by the mutant U2AF35 was suppressed significantly by the co-overexpression of the wild-type protein (Supplementary Fig. 11d), indicating that the effect of the mutant protein was likely to be mediated by inhibition of the functions of the wild-type protein. Given that the NMD pathway, known as mRNA surveillance, provides a post-transcriptional mechanism for recognizing and eliminating abnormal transcripts that prematurely terminate translation²⁷, the result of the GSEA analyses indicated that the mutant U2AF35 induced abnormal RNA splicing in HeLa and TF-1 cells, leading to the generation of unsplined RNA species having a premature stop codon and induction of the NMD activity.

To confirm this, we next performed whole transcriptome analysis in these cells using the GeneChip Human exon 1.0 ST Array (Affymetrix), in which we differentially tracked the behaviour of two discrete sets of probes showing different level of evidence of being exons, that is, 'Core' (authentic exons) and 'non-Core' (more likely introns) sets (Supplementary Methods IV and Supplementary Fig. 12). As shown in Fig. 4d, the Core and non-Core set probes were differentially enriched among probes showing significant difference in expression between wild-type and mutant-transduced cells (false discovery rate (FDR) = 0.01). The Core set probes were significantly enriched in those probes significantly downregulated in mutant U2AF35-transduced cells compared with wild-type U2AF35-transduced cells, whereas the non-Core set probes were enriched in those probes significantly upregulated in mutant U2AF35-transduced cells (Fig. 4e). The significant differential enrichment was also demonstrated, even when all probe sets were included (Fig. 4f). Moreover, the significantly differentially expressed Core set probes tended to be up- and downregulated in wild-type and mutant U2AF35-transduced cells compared with mock-transduced cells, respectively, and vice versa for the differentially expressed non-Core set probes (Fig. 4e). Combined, these exon array results indicated that the wild-type U2AF35 correctly promoted authentic RNA splicing, whereas the mutant U2AF35 inhibited this processes, rendering non-Core and therefore, more likely intronic sequences to remain unsplined.

The abnormal splicing in mutant U2AF35-transduced cells was more directly demonstrated by sequencing mRNAs extracted from HeLa cells, in which expression of the wild-type and mutant (S34F) U2AF35 were induced by doxycycline. First, after adjusting by the total number of mapped reads, the wild-type U2AF35-transduced cells showed an increased read counts in the exon fraction, but reduced counts in other fractions, compared with mutant U2AF35-transduced cells (Fig. 4g). The reads from the mutant-transduced cells were mapped to broader genomic regions compared with those from the wild-type U2AF35-transduced cells, which were largely explained by non-exon reads (Fig. 4h). Finally, the number of those reads that encompassed the authentic exon/intron junctions was significantly increased in mutant U2AF35-transduced cells compared with wild-type U2AF35-transduced cells (Fig. 4i and Supplementary Methods VI). These results clearly demonstrated that failure of splicing ubiquitously occurred in mutant U2AF35-transduced cells. A typical example of abnormal splicing in mutant-transduced cells and the list of significantly unsplined exons are shown in Supplementary Fig. 13 and Supplementary Table 8, respectively.

Biological consequence of U2AF35 mutations

Finally, we examined the biological effects of compromised functions of the E/A splicing complexes. First, TF-1 and HeLa cells were transduced with lentivirus constructs expressing either the S34F U2AF35 mutant or wild-type U2AF35 under a tetracycline-inducible promoter (Fig. 5a and Supplementary Figs 14a and 15a), and cell proliferation was examined after the induction of their expression. Unexpectedly, after the induction of gene expression with

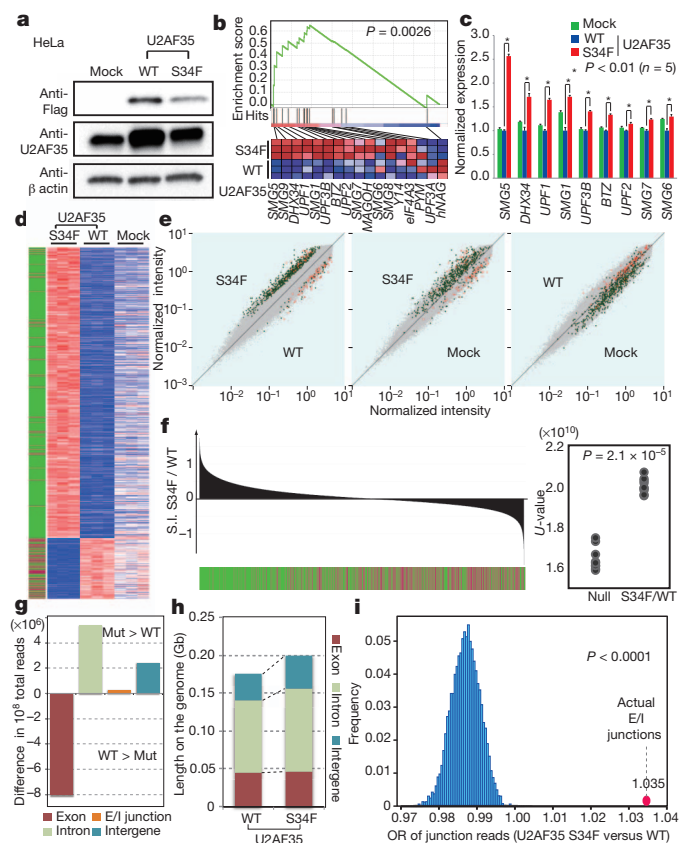


Figure 4 | Altered RNA splicing caused by a U2AF35 mutant. **a**, Western blot analyses showing expression of transduced wild-type or mutant (S34F) U2AF35 in HeLa cells used for the analyses of expression and exon microarrays. **b**, The GSEA demonstrating a significant enrichment of the set of 17 NMD pathway genes among significantly differentially expressed genes between wild-type and mutant U2AF35-transduced HeLa cells. The significance of the gene set was empirically determined by 1,000 gene-set permutations. **c**, The confirmation of the microarray analysis for the expression of nine genes that contributed to the core enrichment in the NMD gene set. Means \pm s.e. are provided for the indicated NMD genes. *P* values were determined by the Mann–Whitney *U* test. **d**, Significantly upregulated and downregulated probe sets (FDR = 0.01) in mutant U2AF35-transduced cells compared with wild-type U2AF35-transduced cells in triplicate exon array experiments are shown in a heat map. The origin of each probe set is depicted in the left lane, where red and green bars indicate the Core and non-Core sets, respectively. **e**, Pair-wise scatter plots of the normalized intensities of entire probe sets (grey) across different experiments. The Core and non-Core set probes that were significantly differentially expressed between the wild-type and mutant U2AF35-transduced cells are plotted in red and green, respectively. **f**, Distribution of the Core (red) and non-Core (green) probe sets within the entire probe sets ordered by splicing index (S.I.; Supplementary Methods IV), calculated between wild-type and mutant U2AF35-transduced cells. In the right panel, the differential enrichment of both probe sets was confirmed by Mann–Whitney *U* test. **g**, Difference in read counts for the indicated fractions per 10^8 total reads in RNA sequencing between wild-type and mutant U2AF35-expressing HeLa cells analysis. Increased/decreased read counts in mutant U2AF35-expressing cells are plotted upward/downward, respectively. **h**, Comparison of the genome coverage by the indicated fractions in wild-type- and mutant-U2AF35-expressing cells. The genome coverage was calculated for each fraction within the 10^8 reads randomly selected from the total reads and averaged for ten independent selections. **i**, The odds ratio of the junction reads within the total mapped reads was calculated between the two experiments (red circle), which was evaluated against the 10,000 simulated values under the null hypothesis (histogram in blue).

doxycycline, the mutant U2AF35-transduced cells, but not the wild-type U2AF35-transduced cells, showed reduced cell proliferation (Fig. 5b and Supplementary Fig. 15b) with a marked increase in the G2/M fraction (G2/M arrest) together with enhanced apoptosis as

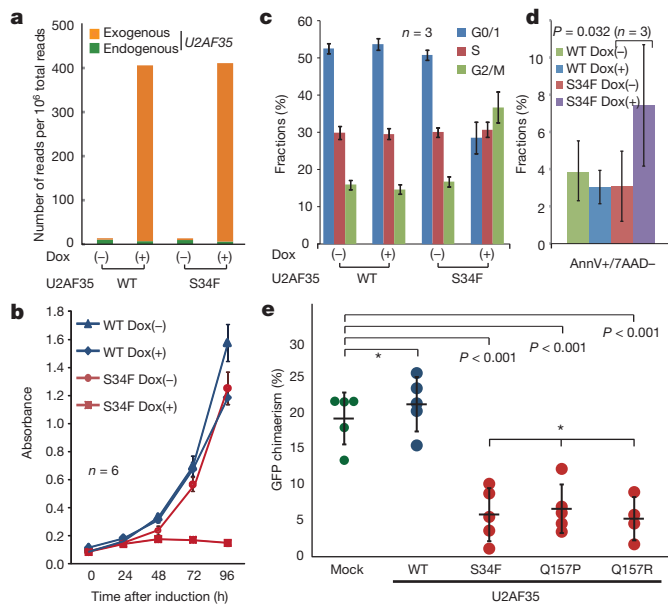


Figure 5 | Functional analysis of mutant U2AF35. **a**, Expression of endogenous and exogenous U2AF35 transcripts in HeLa cells before and after induction determined by RNA sequencing. U2AF35 transcripts were differentially enumerated for endogenous and exogenous species, which were discriminated by the Flag sequence. **b**, Cell proliferation assays of U2AF35-transduced HeLa cells, where cell numbers were measured using cell-counting apparatus and are plotted as mean absorbance \pm s.d. **c**, The flow cytometry analysis of propidium iodide (PI)-stained HeLa cells transduced with the different U2AF35 constructs. Mean fractions \pm s.d. in G0/G1, S and G2/M populations after the induction of U2AF35 expression are plotted. **d**, Fractions of the annexin V-positive (AnnV+) populations among the 7-amino-actinomycin D (7AAD)-negative population before and after the induction of U2AF35 expression are plotted as mean \pm s.d. for indicated samples. The significance of difference was determined by paired *t*-test. **e**, Competitive reconstitution assays for CD34-negative KSL cells transduced with indicated U2AF35 mutants. Chimaerism in the peripheral blood 6 weeks after transplantation are plotted as mean %EGFP-positive Ly5.1 cells \pm s.d., where outliers were excluded from the analysis. The significance of differences was evaluated by the Grubbs test with Bonferroni's correction for multiple testing. *not significant.

indicated by the increased sub-G1 fraction and annexin V-positive cells (Fig. 5c, d, Supplementary Fig. 14b and Supplementary Methods VI). To confirm the growth-suppressive effect of U2AF35 mutants *in vitro*, a highly purified haematopoietic stem cell population (CD34⁺c-Kit⁺Scal⁺Lin⁻, CD34⁺KSL) prepared from C57BL/6 (B6)-Ly5.1 mouse bone marrow²⁸ was retrovirally transduced with either the mutant (S34F, Q157P and Q157R) or wild-type U2AF35, or the mock constructs, each harbouring the EGFP marker gene (Supplementary Fig. 16). The ability of these transduced cells to reconstitute the haematopoietic system was tested in a competitive reconstitution assay. The transduced cells were mixed with whole bone marrow cells from B6-Ly5.1/5.2 F1 mice, transplanted into lethally irradiated B6-Ly5.2 recipients, and peripheral blood chimaerism derived from EGFP-positive cells was assessed 6 weeks after transplantation by flow cytometry. We confirmed that each recipient mouse received comparable numbers of EGFP-positive cells among the different retrovirus groups by estimating the percentage of EGFP-positive cells and overall proliferation in transduced cells by *ex vivo* tracking. Also no significant difference was observed in their homing capacity to bone marrow as assessed by transwell migration assays (Supplementary Fig. 17). As shown in Fig. 5e, the wild-type U2AF35-transduced cells showed a slightly higher reconstitution capacity than the mock-transduced cells. On the other hand, the recipients of the cells transduced with the various U2AF35 mutants showed significantly lower EGFP-positive cell chimaerism than those of either the mock- or the wild-type U2AF35-transduced

cells, indicating a compromised reconstitution capacity of the haematopoietic stem/progenitor cells expressing the U2AF35 mutants. In summary, these mutants lead to loss-of-function of U2AF35 most probably by acting in a dominant-negative fashion to the wild-type protein.

Discussion

Our whole-exome sequencing study unexpectedly unmasked a complexity of novel pathway mutations found in approximately 45% to 85% of myelodysplasia patients depending on the disease subtypes, which affected multiple but distinctive components of the splicing machinery and, as such, demonstrated the unquestionable power of massively parallel sequencing technologies in cancer research.

The RNA splicing system comprises essential cellular machinery, through which eukaryotes can achieve successful transcription and guarantee the functional diversity of their protein species using alternative splicing in the face of a limited number of genes²⁹. Accordingly, the meticulous regulation of this machinery should be indispensable for the maintenance of cellular homeostasis³⁰, deregulation of which causes severe developmental abnormalities^{31,32}. The current discovery of frequent mutations of the splicing pathway in myelodysplasia, therefore, represents another remarkable example that illustrates how cancer develops by targeting critical cellular functions. It also provides an intriguing insight into the mechanism of 'cancer specific' alternative splicing, which have long been implicated in the development of cancer, including MDS and other haematopoietic neoplasms^{33,34}.

In myelodysplasia, the major targets of spliceosome mutations seemed to be largely confined to the components of the E/A splicing complex, among others to SF3B1, SRSF2, U2AF35 and ZRSR2, and to a lesser extent, to SF3A1, SF1, U2AF65 and PRPF40B. The broad coverage of the wide spectrum of spliceosome components in our exome sequencing was likely to preclude frequent involvement of other components on this pathway (Supplementary Fig. 18). The surprising frequency and specificity of these mutations in this complex, together with the mutually exclusive manner they occurred, unequivocally indicate that the compromised function of the E/A complex is a hallmark of this unique category of myeloid neoplasms, playing a central role in the pathogenesis of myelodysplasia. The close relationship between the mutation types and unique disease subtypes also support their pivotal roles in MDS.

Given the critical functions of the E/A splicing complex on the precise 3'SS recognition, the logical consequence of these relevant mutations would be the impaired splicing involving diverse RNA species. In fact, when expressed in HeLa cells, the mutant U2AF35 induced global abnormalities of RNA splicing, leading to increased production of transcripts having unspliced intronic sequences. On the other hand, the functional link between the abnormal splicing of RNA species and the phenotype of myelodysplasia is still unclear. Mutant U2AF35 seemed to suppress cell growth/proliferation and induce apoptosis rather than confer a growth advantage or promote clonal selection. ZRSR2 knockdown in HeLa cells has been reported to also result in reduced viability, arguing for the common consequence of these pathway mutations³⁵. These observations suggested that the oncogenic actions of these splicing pathway mutations are distinct from what is expected for classical oncogenes, such as mutated kinases and signal transducers, but could be more related to cell differentiation. Of note in this regard, the commonest clinical presentation of MDS is severe cytopenia in multiple cell lineages due to ineffective haematopoiesis with increased apoptosis rather than unlimited cell proliferation¹. In this regard, lessons may be learned from the recent findings on the pathogenesis of the 5q- syndrome, where haploinsufficiency of RPS14 leads to increased apoptosis of erythroid progenitors, but not myeloproliferation^{36,37}.

A lot of issues remain to be answered, however, to establish the functional link between these splicing pathway mutations and the

pathogenesis of MDS, where the broad spectrum of RNA species affected by impaired splicing hampers identification of responsible gene targets. Moreover, the mutated components of the splicing machinery have distinct function of their own other than direct regulation of RNA splicing, involved in elongation and DNA stability, which may be important to determine specific disease phenotypes. Clearly, more studies are required to answer these questions through understanding of the molecular basis of their oncogenic actions.

METHODS SUMMARY

Whole-exome sequencing of paired tumour/normal DNA samples from the 29 patients was performed after informed consent was obtained. SNP array-based copy number analysis was performed as previously described^{17,18}. Mutation analysis of the splicing pathway genes in a set of 582 myeloid neoplasms were performed by first screening mutations in PCR-amplified pooled targets from 12 individuals, followed by validation/identification of the candidate mutations within the corresponding 12 individuals by Sanger sequencing. Flag-tagged cDNAs of the wild-type and mutant *U2AF35* were generated by *in vitro* mutagenesis, constructed into a murine stem cell virus-based retroviral vector as well as a tetracycline-inducible lentivirus-based expression vector, and used for gene transfer to CD34⁺KSL cells and cultured cell lines, with EGFP marking, respectively. Total RNA was extracted from wild-type or mutant *U2AF35*-transduced HeLa and TF-1 cells, and analysed on microarrays. RNA sequencing was performed according to the manufacturer's instructions (Illumina). Cell proliferation assays (MTT assays) on HeLa and TF-1 cells stably transduced with lentivirus *U2AF35* constructs were performed in the presence or absence of doxycycline. For competitive reconstitution assays, CD34⁺KSL cells collected from C57BL/6 (B6)-Ly5.1 mice were retrovirally transduced with various *U2AF35* constructs with EGFP marking, and transplanted with competitor cells (B6-Ly5.1/5.2 F1 mouse origin) into lethally irradiated B6-Ly5.2 mice 48 h after gene transduction. Frequency of EGFP-positive cells was assessed in peripheral blood by flow cytometry 6 weeks after the transplantation (Supplementary Methods VII). The primer sets used for validation of gene mutations and qPCR of NMD gene expression are listed in Supplementary Tables 9–11. A complete description of the materials and methods is provided in the Supplementary Information. This study was approved by the ethics boards of the University of Tokyo, Munich Leukaemia Laboratory, University Hospital Mannheim, University of Tsukuba, Tokyo Metropolitan Ohtsuka Hospital and Chang Gung Memorial Hospital. Animal experiments were performed with approval of the Animal Experiment Committee of the University of Tokyo.

Received 7 June; accepted 24 August 2011.

Published online 11 September 2011.

- Corey, S. J. *et al.* Myelodysplastic syndromes: the complexity of stem-cell diseases. *Nature Rev. Cancer* **7**, 118–129 (2007).
- Ma, X., Does, M., Raza, A. & Mayne, S. T. Myelodysplastic syndromes: incidence and survival in the United States. *Cancer* **109**, 1536–1542 (2007).
- Bejar, R., Levine, R. & Ebert, B. L. Unraveling the molecular pathophysiology of myelodysplastic syndromes. *J. Clin. Oncol.* **29**, 504–515 (2011).
- Sanada, M. *et al.* Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. *Nature* **460**, 904–908 (2009).
- Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40**, 722–729 (2008).
- Chapman, M. A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467–472 (2011).
- Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465**, 473–477 (2010).
- Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
- Metzker, M. L. Sequencing technologies — the next generation. *Nature Rev. Genet.* **11**, 31–46 (2010).
- Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nature Biotechnol.* **26**, 1135–1145 (2008).
- Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
- Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* **469**, 539–542 (2011).
- Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).
- Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
- Yan, X. J. *et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene *DNMT3A* in acute monocytic leukemia. *Nature Genet.* **43**, 309–315 (2011).

- Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101–105 (2011).
- Nannay, Y. *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res.* **65**, 6071–6079 (2005).
- Yamamoto, G. *et al.* Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of Affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am. J. Hum. Genet.* **81**, 114–126 (2007).
- Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
- Tronchère, H., Wang, J. & Fu, X. D. A protein related to splicing factor *U2AF³⁵* that interacts with *U2AF⁵⁵* and SR proteins in splicing of pre-mRNA. *Nature* **388**, 397–400 (1997).
- Bevilacqua, L. *et al.* A population-specific *HTR2B* stop codon predisposes to severe impulsivity. *Nature* **468**, 1061–1066 (2010).
- Calvo, S. E. *et al.* High-throughput, pooled sequencing identifies mutations in *NUBPL* and *FOXRED1* in human complex I deficiency. *Nature Genet.* **42**, 851–858 (2010).
- Haase, D. *et al.* New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. *Blood* **110**, 4385–4395 (2007).
- Xiao, R. *et al.* Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. *Mol. Cell. Biol.* **27**, 5393–5402 (2007).
- Morin, R. D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nature Genet.* **42**, 181–185 (2010).
- Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
- Maquat, L. E. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nature Rev. Mol. Cell Biol.* **5**, 89–99 (2004).
- Ema, H. *et al.* Adult mouse hematopoietic stem cells: purification and single-cell assays. *Nature Protocols* **1**, 2979–2987 (2007).
- Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Rev. Mol. Cell Biol.* **10**, 741–754 (2009).
- Ni, J. Z. *et al.* Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21**, 708–718 (2007).
- He, H. *et al.* Mutations in *U4atac* snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. *Science* **332**, 238–240 (2011).
- Edery, P. *et al.* Association of TALS developmental disorder with defect in minor splicing component *U4atac* snRNA. *Science* **332**, 240–243 (2011).
- David, C. J. & Manley, J. L. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* **24**, 2343–2364 (2010).
- Pajares, M. J. *et al.* Alternative splicing: an emerging topic in molecular and clinical oncology. *Lancet Oncol.* **8**, 349–357 (2007).
- Shen, H., Zheng, X., Luecke, S. & Green, M. R. The *U2AF35*-related protein Urp contacts the 3' splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. *Genes Dev.* **24**, 2389–2394 (2010).
- Barlow, J. L. *et al.* A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q- syndrome. *Nature Med.* **16**, 59–66 (2010).
- Ebert, B. L. *et al.* Identification of *RPS14* as a 5q- syndrome gene by RNA interference screen. *Nature* **451**, 335–339 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by Grant-in-Aids from the Ministry of Health, Labor and Welfare of Japan and from the Ministry of Education, Culture, Sports, Science and Technology, and also by the Japan Society for the Promotion of Science (JSPS) through the 'Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)', initiated by the Council for Science and Technology Policy (CSTP). pGCDNsamiRESEGFP vector was a gift from M. Onodera. We thank Y. Mori, O. Hagiwara, M. Nakamura and N. Mizota for their technical assistance. We are also grateful to K. Ikeuchi and M. Ueda for their continuous encouragement throughout the study.

Author Contributions Y.Sh., Y.Sa., A.S.-O., Y.N., M.N., G.C., R.K. and S.Miyano were committed to bioinformatics analyses of resequencing data. M.Sa., A.S.-O. and Y.Sa. performed microarray experiments and their analyses. R.Y., T.Y., M.O., M.Sa., A.K., M.Sh. and H.N. were involved in the functional analyses of *U2AF35* mutants. N.O., M.S.-Y., K.I., H.M., W.-K.H., F.N., D.N., T.H., C.H., S.Miyawaki, S.C., H.P.K. and L.-Y.S. collected specimens and were also involved in planning the project. K.Y., Y.N., Y.Su., A.S.-O. and S.S. processed and analysed genetic materials, library preparation and sequencing. K.Y., M.Sa., Y.Sh., A.S.-O., Y. Sa. and S.O. generated figures and tables. S.O. led the entire project and wrote the manuscript. All authors participated in the discussion and interpretation of the data and the results.

Author Information Sequence data have been deposited in the DDBJ repository under accession number DRA000433. Microarray data have been deposited in the GEO database under accession numbers GSE31174 (for SNP arrays), GSE31171 (for exon arrays) and GSE31172 (for expression arrays). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.O. (sogawa-tyk@umin.ac.jp).

Diffraction-unlimited all-optical imaging and writing with a photochromic GFP

Tim Grotjohann^{1*}, Ilaria Testa^{1*}, Marcel Leutenegger^{1*}, Hannes Bock¹, Nicolai T. Urban¹, Flavie Lavoie-Cardinal¹, Katrin I. Willig¹, Christian Eggeling¹, Stefan Jakobs^{1,2} & Stefan W. Hell¹

Lens-based optical microscopy failed to discern fluorescent features closer than 200 nm for decades, but the recent breaking of the diffraction resolution barrier by sequentially switching the fluorescence capability of adjacent features on and off is making nanoscale imaging routine. Reported fluorescence nanoscopy variants switch these features either with intense beams at defined positions or randomly, molecule by molecule. Here we demonstrate an optical nanoscopy that records raw data images from living cells and tissues with low levels of light. This advance has been facilitated by the generation of reversibly switchable enhanced green fluorescent protein (rsEGFP), a fluorescent protein that can be reversibly photoswitched more than a thousand times. Distributions of functional rsEGFP-fusion proteins in living bacteria and mammalian cells are imaged at <40-nanometre resolution. Dendritic spines in living brain slices are super-resolved with about a million times lower light intensities than before. The reversible switching also enables all-optical writing of features with subdiffraction size and spacings, which can be used for data storage.

In a fluorescence microscope, diffraction prevents (excitation) light being focused more sharply than $\lambda/(2\text{NA})$, with λ being the wavelength of light and NA the numerical aperture of the lens. Thus, as they are illuminated together, features residing any closer together than this distance also fluoresce together and appear in the image as a single blur. The diffraction resolution barrier can be overcome by forcing such nearby features to fluoresce sequentially, but this strategy clearly requires a mechanism for keeping fluorophores that are exposed to excitation light non-fluorescent^{1–3}.

In stimulated emission depletion (STED) microscopy^{1,4}, this is accomplished by the so-called STED beam, which turns the fluorescence capability of fluorophores off by a photon-induced de-excitation. Because at least a single de-exciting photon must be available within the lifetime ($\tau \approx 1\text{--}5\text{ ns}$) of the fluorescent molecular state, the intensity of the focal STED beam must exceed the threshold $I_s = C\tau^{-1}$ with C accounting for the probability of a STED beam photon to interact with the fluorophore^{1,4}. The STED beam, usually formed as a doughnut overlaid with the excitation beam, features a central point of zero intensity at which the fluorophores can still assume the fluorescent state. As this point can be positioned with arbitrary precision in space, the coordinate of the emitting (on-state) fluorophores is known at any instant: it is the position of zero intensity^{3,5,6} and its immediate vicinity, where the STED beam is still weaker than I_s . The diameter of this area is given by $d \approx \lambda/[2\text{NA} \times (1 + I_m/I_s)^{1/2}]$, with I_m (typically $\gg I_s$) denoting the intensity at the doughnut crest. Hence, features that are (just slightly) more apart than $d \ll \lambda/(2\text{NA})$ cannot fluoresce at the same time even when simultaneously illuminated by excitation light⁶. Scanning the beams across the sample and recording the fluorescence yields images of subdiffraction resolution d automatically and irrespective of the fluorophore concentration in the sample.

De-excitation by stimulated emission is the most basic and general mechanism for modulating the fluorescence ability of a molecule. However, by requiring light intensities $> I_s \approx 1\text{--}10\text{ MW cm}^{-2}$, attaining high resolutions by this mechanism necessitates large I_m values. For

example, $d < 40\text{ nm}$ typically entails $I_m = 100\text{--}500\text{ MW cm}^{-2}$ (ref. 6). Although intensities of this order have been demonstrated to be live-cell compatible^{4,7–10}, all-optical nanoscopy methods operating at fundamentally lower light levels are highly in demand^{2,5,11–13}, because they allow larger fields of view^{5,14} and can avoid photodamage.

A route to low light level operation is to replace STED with a fluorescence switching mechanism having a lower threshold I_s (refs 2, 5, 11–13). Following the equation for I_s , this can be realized by exploiting transitions between fluorophore states of longer lifetime $\tau \gg 1\text{ }\mu\text{s}$ (refs 2, 5, 11). Hence, it has been suggested that fluorescence can be switched by transferring the fluorophores transiently to a generic metastable dark (triplet) state of $\tau \approx 10^{-3}\text{--}100\text{ ms}$ (refs 2, 15). A more attractive option is to use fluorophores that can be explicitly ‘photoswitched’^{5,11}, for example, by photoisomerization. Hence, in 2003 it was proposed to implement a STED-like microscope with STED being replaced by a reversible on–off switch as encountered in organic photochromic fluorophores and reversibly photoswitchable fluorescent proteins (RSFPs)^{5,11}.

In fact, this strategy is even more general because any reversible transition between a signalling and a non-signalling state can be used for breaking the diffraction barrier¹⁵. Therefore, all concepts that switch the fluorescence capability of molecules at sample coordinates predefined by patterns of light have been generalized under the name RESOLFT^{6,12}, which stands for reversible saturable optical (fluorescence) transition between two states. Note that a photoswitch is a perfect saturable transition. Concomitantly, the concept was extended to subdiffraction writing^{11,13} and data storage, in which case the on-state is a reactive state from which the molecule can be made permanent whereas the off-state serves as a temporary ‘mask’ defining the structure to be written.

Super-resolution by switching RSFPs was shown in 2005¹⁶, but this study relied on asFP595¹⁷, a tetrameric protein with low fluorescence quantum yield. Moreover, when translating the light pattern across the sample, the proteins faded after a few cycles, implying that features that had been turned off could not be turned on again in order to be

¹Department of NanoBiophotonics, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. ²University of Göttingen Medical School, Robert-Koch-Str. 40, 37075 Göttingen, Germany.

*These authors contributed equally to this work.

read out. Biological imaging therefore remained unviable¹⁸. Other studies using a variant of the RSFP called dronpa¹⁹ faced the same challenge²⁰. As a rule of thumb, an m -fold resolution improvement along a certain direction requires $\sim m$ switching cycles, meaning that $m = 10$ along the x - and y -axes entails $\sim m^2 = 100$ cycles, whereas $\sim 1,000$ cycles are required for x , y and z (ref. 6). Thus, for RESOLFT super-resolution, the number of switching cycles afforded by the fluorophore assumes a vital role.

Because they are able to generate an image with a single on-off cycle^{3,6,21}, the super-resolution concepts called (F)PALM^{22,23} and STORM^{21,24}, which have emerged in the interim, have successfully harnessed the switching between metastable states for gaining sub-diffraction resolution. However, these methods rely on the imaging and computation-aided localization of individual fluorophores amidst the scattering and autofluorescence background common in (living) cells and tissues. Moreover, rapid localization of a sufficiently large number of fluorophores requires the excitation light to be intense^{22,23}. In contrast, a RESOLFT approach is able to instantly record the emission from all fluorophores attached to the nanosized feature of interest⁶, and can be easily combined with confocal microscopy for three-dimensional imaging and background suppression. Yet again, because all RSFPs, conventional fluorescent proteins²⁵ and photochromic rhodamines²⁶ seemed unsuitable (Supplementary Fig. 1), an all-optical nanoscopy approach operating at low light levels appeared unviable.

Similarly, although STED/RESOLFT-inspired optical writing with photochromic compounds has been shown to yield structures $< \lambda/(2\text{NA})$, writing such structures with spacings $< \lambda/(2\text{NA})$ remained challenging^{27–30}, again the impediment being the requirement of many on-off cycles before the structure is made permanent. Here we introduce a RSFP enabling both low-light-level all-optical nanoscopy of living cells and tissues, and far-field optical writing and reading of patterns of subdiffraction size and density.

Generating a reversibly switchable GFP

All fluorescent proteins have a similar fold, namely an 11-stranded β -barrel with a central helix containing the chromophore, which is typically in a *cis*-configuration³¹. Light-driven switching of RSFPs generally involves an isomerization of the chromophore, frequently coupled with a change of its protonation state^{32–36}. We started from EGFP³⁷ and identified, using its X-ray structure³¹, amino acid residues the exchange of which was expected to facilitate isomerization. We expressed numerous EGFP variants in *Escherichia coli* and screened for colonies expressing an RSFP with an automated microscope. To this end, we alternated site-directed and error-prone mutagenesis while maintaining the key amino acids of EGFP (that is, F64L and S65T)³⁷; we concomitantly introduced A206K to ensure that the protein remained a monomer³⁸.

The amino acid exchange Q69L was sufficient to make EGFP(A206K) reversibly switchable, but the resulting on-off contrast was low. Although it makes the protein switchable³⁹, we avoided the mutation E222Q because it seemed to reduce the number of cycles. After analysing $\sim 30,000$ clones, we identified EGFP(Q69L/V150A/V163S/S205N/A206K) (Supplementary Fig. 2) that could be reversibly switched on at $\lambda = 405$ nm and off at 491 nm, and named it reversibly switchable EGFP (rsEGFP).

At equilibrium, rsEGFP adopts a bright on-state (fluorescence quantum yield $\Phi_{\text{FL}} = 0.36$; extinction coefficient $\epsilon = 47,000 \text{ M}^{-1} \text{ cm}^{-1}$ (Supplementary Table 1)). In the on-state, rsEGFP exhibits a single absorption band peaking at 491 nm (Fig. 1a), corresponding to the ionized state of the phenolic hydroxyl of the chromophore⁴⁰. The pK_a of the chromophore is 6.5 (Supplementary Fig. 3). Absorption at 490 nm yields fluorescence peaking at 510 nm and, in a competing process, switches rsEGFP off (Figs 1a–c). Prolonged irradiation of a pH 7.5 solution of purified rsEGFP at ~ 490 nm reduces the rsEGFP fluorescence to 1–2% of its initial value. The off-state exhibits a single absorption band

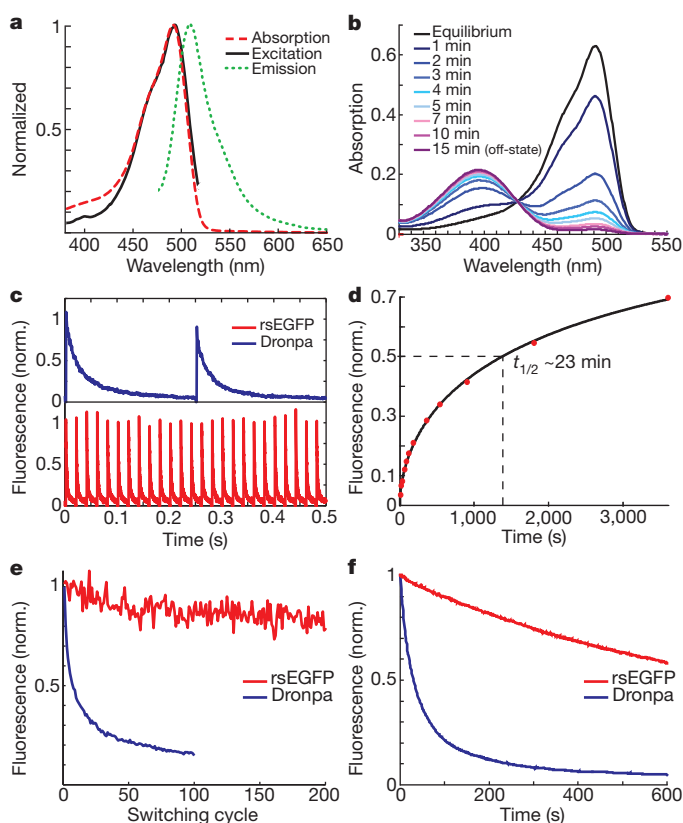


Figure 1 | Properties of rsEGFP. **a**, Absorption (red dashed line), excitation (solid black line) and fluorescence (dotted green line) spectrum of rsEGFP in the fluorescent equilibrium state at pH 7.5. **b**, Absorption spectra obtained at different time points during irradiation with 488-nm light. **c**, Switching curves of dronpa (blue) and rsEGFP (red) immobilized in PAA using the same intensities. Switching was performed by alternating irradiation at 405 nm (2 kW cm^{-2}) and at 491 nm (0.6 kW cm^{-2}). The duration of off-switching at 491 nm was chosen such that the fluorescence reached a minimum; irradiation with 405 nm was chosen so that the proteins were fully switched. **d**, Relaxation of rsEGFP embedded in PAA from the off-state into the fluorescent equilibrium state at 22°C . The black line is a stretched exponential fit with a stretching factor of ~ 0.6 accounting for inhomogeneous spectral broadening or the involvement of multiple dark states. **e**, Fluorescence per switching cycle normalized to the initial fluorescence, with the same light intensities and switching durations as in **c**. **f**, Photobleaching: rsEGFP and dronpa embedded in a PAA layer were kept in their on-states by continuous irradiation at 405 nm (1 kW cm^{-2}), while fluorescence was probed by irradiation at 491 nm (3 kW cm^{-2}).

at 396 nm, corresponding to the neutral state of the chromophore (Fig. 1b). Excitation at this band switches the protein back to the on-state. At room temperature rsEGFP converts spontaneously from the off- into the on-state with a half-time of ~ 23 min (Fig. 1d).

We compared the properties of rsEGFP with that of the well-known RSFP dronpa¹⁹. With the proteins embedded in a 12.5% polyacrylamide (PAA) layer and using light of 491 nm (0.6 kW cm^{-2}) and 405 nm (2 kW cm^{-2}), a complete on-off cycle took 250 ms for dronpa and 20 ms for rsEGFP (Fig. 1c). Dronpa went through < 10 cycles before its fluorescence was reduced to 50%, whereas rsEGFP went through $\sim 1,200$ cycles under the same conditions (Fig. 1e). To compare bleaching, dronpa and rsEGFP were kept in the on-state by continuous irradiation at 405 nm (1 kW cm^{-2}) while fluorescence was generated by irradiation at 491 nm (3 kW cm^{-2}). Whereas dronpa fluorescence was reduced to 50% within $t_{1/2} \approx 30$ s, for rsEGFP we measured $t_{1/2} \approx 800$ s (Fig. 1f). The rsEGFP chromophore matured with a half-time of ~ 3 h at 37°C (Supplementary Fig. 4). The protein behaved as a monomer *in vitro* (Supplementary Fig. 5), could be fused to various proteins, including α -tubulin and histone H2B (Supplementary Fig. 6), and was repeatedly switchable in living cells (Supplementary Fig. 7).

Rewritable data storage

To analyse whether immobilized rsEGFP could be used for repeated short-term data storage⁴¹, we coated a microscope slide with a $<1\text{-}\mu\text{m}$ thin layer of rsEGFP ($\sim 0.03\text{ mM}$) in PAA. Switching and reading by illumination at 405 nm and 491 nm in a scanning confocal set-up provided an on-off contrast of $\sim 50:1$. We translated the text of 25 Grimm's fairy stories (<http://www.gutenberg.org/files/11027/11027.txt>) into 7-bit binary ASCII code ('0': off; '1': on) and wrote and read the $\sim 270,000$ letters into a $17\text{ }\mu\text{m} \times 17\text{ }\mu\text{m}$ region in 6,596 frames, each comprising 41 letters (287 bits) (Fig. 2). Individual bits were $\sim 0.5\text{ }\mu\text{m}$ in diameter with $1\text{ }\mu\text{m}$ centre-to-centre spacing, corresponding to a DVD storage density. Discriminating '0' from '1' by a simple threshold entailed 7 bit errors within the entire data set. After $\sim 6,600$ read/write cycles in the same region, the average fluorescence of the '1' was reduced by $\sim 35\%$ (Supplementary Fig. 8). Hence, the same rsEGFP layer can be used for $\sim 15,000$ read/write processes.

RESOLFT nanoscopy of living samples

Next, we implemented a scanning confocal set-up with a 405 nm (ultraviolet) beam for switching the rsEGFP on, a 491 nm (blue) beam for eliciting fluorescence, and a doughnut-shaped 491 nm beam for the off-switching (Supplementary Fig. 9). We fused rsEGFP to the amino-terminus of the bacterial actin homologue MreB⁴² and expressed the fusion protein in *E. coli* bacteria. Living bacteria on agar-coated slides were recorded by first irradiating each pixel for $100\text{ }\mu\text{s}$ with ultraviolet light (1 kW cm^{-2}), thus activating most of the rsEGFP in the focal volume. Then the doughnut-shaped blue beam ($I_m \approx 1\text{ kW cm}^{-2}$) was applied for 10–20 ms to switch all the rsEGFP molecules off, except those located within $d/2$ distance from the doughnut centre. Lastly, the rsEGFP fluorescence was read out for 1–2 ms by the 491-nm beam ($\sim 1\text{ kW cm}^{-2}$). The sequence was repeated for each sample pixel.

The double-helical cytoskeletal structure of rsEGFP–MreB is more clearly revealed by RESOLFT than by its confocal counterpart

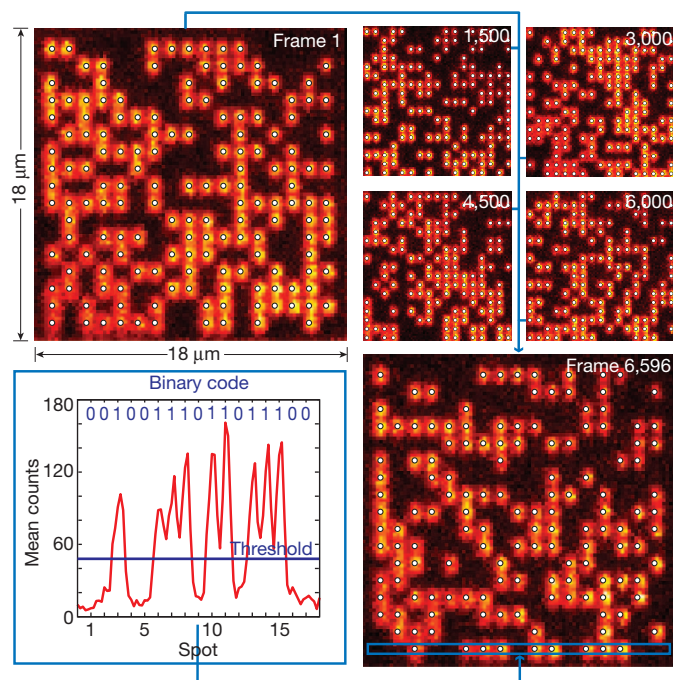


Figure 2 | Rewritable data storage. The text of 25 Grimm's fairy stories (ASCII code; 1.9 Mbits) consecutively written and read on a $17 \times 17\text{ }\mu\text{m}$ area of a PAA layer containing rsEGFP, with bits written as spots (representative frames shown). The white dots mark spots that were recognized as set bits ('1's). The graph shows an intensity profile along the indicated area, averaged over three pixels along the y-axis. The blue line indicates the threshold used to assign read spots to '0's or '1's.

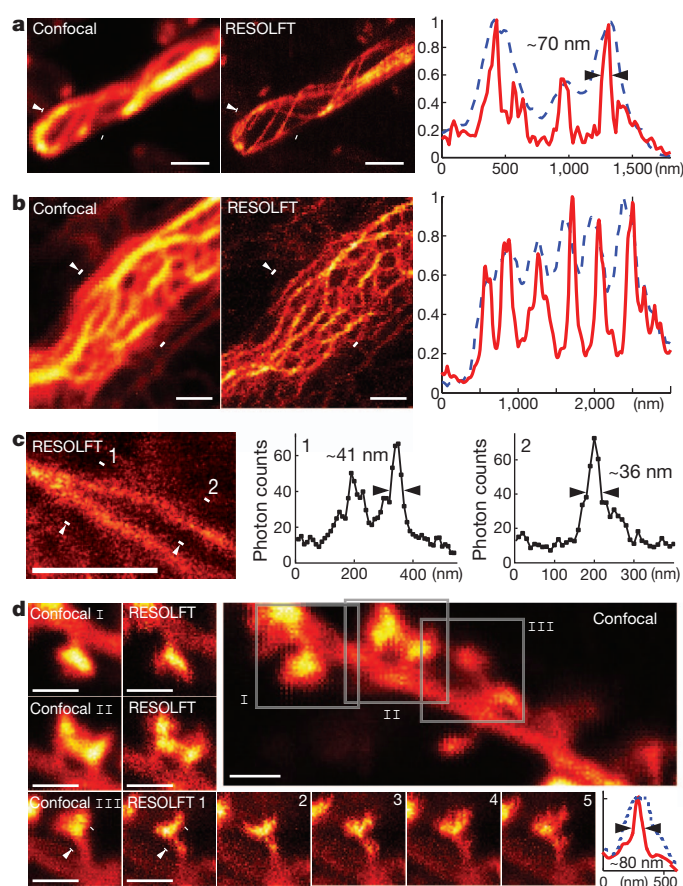


Figure 3 | RESOLFT nanoscopy of living cells. **a**, *E. coli* bacterium expressing rsEGFP–MreB: confocal (left) and corresponding RESOLFT (middle) image. **b**, Mammalian (PtK2) cell expressing keratin-19–rsEGFP imaged in the confocal (left) and the RESOLFT (middle) mode. **a**, **b**, Graphs show the normalized fluorescence profiles between the two white markers with the white arrowhead indicating the direction (solid red, RESOLFT; dashed blue, confocal). **c**, RESOLFT image (left) of keratin-19–rsEGFP filaments in a PtK2 cell recorded with a pixel size of $10\text{ nm} \times 10\text{ nm}$; smoothed with a low-pass Gaussian filter of 1.2 pixel width. Graphs 1 and 2 extracted from the image as indicated reveal resolution $d < 40\text{ nm}$. **d**, Dendrite within a living organotypic hippocampal slice expressing lifeact–rsEGFP. Main image: confocal overview. I–III: three spines, as indicated on the main image, each imaged in the confocal (left) and the RESOLFT mode (right). Spine III was repeatedly imaged in the RESOLFT mode within 5 min, demonstrating the changes over time. Graph: normalized profile across a spine neck as imaged in the RESOLFT (solid red) or the confocal mode (dashed blue) between the two white markers. Scale bars, $1\text{ }\mu\text{m}$.

(Fig. 3a). The RESOLFT image of a typical filament showed a full-width half-maximum (FWHM) of $\sim 70\text{ nm}$. Because this value seemed to be determined by the thickness of the filament itself, a more accurate upper limit for the resolution d is obtained by imaging the finer keratin-19–rsEGFP intermediate filament network in living mammalian cells (Fig. 3b, c). Line profiles from recorded data gave $d < 40\text{ nm}$ corresponding to a 5–6-fold all-optical resolution improvement over confocal microscopy (Fig. 3c).

To investigate its applicability to living brain tissue, we locally injected viral particles carrying a lifeact–rsEGFP construct into a cultured organotypic hippocampal brain slice. Lifeact is a 17-amino-acid-long peptide with high affinity to filamentous actin⁴³. RESOLFT revealed fine morphological differences between the spines protruding from a dendrite (Fig. 3d). A profile through a spine neck showed a FWHM of $< 80\text{ nm}$. Electron microscopy of similar samples demonstrated that this value is close to the actual size of the spine necks themselves⁴⁴, suggesting a resolution d substantially $< 80\text{ nm}$. Repeated imaging revealed dynamic changes over 5 min (Fig. 3d). Altogether, the resolution is comparable to that provided by STED

on similar structures¹⁰, but here it is obtained with light intensities lower by about a million times.

RESOLFT optical data storage

For investigating subdiffraction resolution writing, an rsEGFP layer was prepared as previously outlined. The writing entailed (1) an ultraviolet beam (405 nm, 1 kW cm^{-2}) applied for 100 μs to switch rsEGFP on, (2) a 2-ms break for equilibration, (3) a doughnut-shaped blue beam (491 nm, 0.5 kW cm^{-2}) lasting 20 ms confining the on-state within $d/2$ around the doughnut centre, and (4) an ~ 2 -ms 532 nm beam ($\sim 900 \text{ kW cm}^{-2}$) for transferring on-state rsEGFP to a permanent off (bleached) state (Fig. 4a) (Supplementary Fig. 10a). Lastly, the rsEGFP molecules located outside this region were switched back on, which is critical for writing another feature within subdiffraction proximity.

We wrote nine patterns of 3×3 bit fields in an rsEGFP layer, with 250 nm centre-to-centre separation between individual bits (Fig. 4b), both in the conventional and in the RESOLFT mode. Whereas conventional writing and/or confocal reading blurred the data, the bits were fully discernible when both writing and reading were performed by RESOLFT. We wrote and read the data down to distances of 200 nm between the individual bits (Supplementary Fig. 10b). Hence this scheme allowed storing and reading out bits ~ 4 times more densely than by regular focusing. The structures could be read 5–10 times.

Discussion and conclusion

The many-switching cycles afforded by the fluorescence protein rsEGFP reported here has facilitated live-cell RESOLFT microscopy, a super-resolution microscopy that is similar to STED microscopy in usability but operates at $\sim 10^6$ times lower levels of light. Multiphoton-induced optical damage⁴⁵ can therefore be virtually excluded. The fundamental reduction in optical intensity required for the on–off

switching stems from the fact that the fluorescence capability of the molecule is not modulated by disallowing the population of its nano-second fluorescent state, but rather by toggling it between two long-lived ground states, one in which the fluorophore remains dark when exposed to the excitation light.

RESOLFT is readily combined with confocal imaging, which increases its use in scattering living samples. In fact, the imaging of neuronal spines in living organotypical brain slices testifies this potential. Although the recording time reported here is still of the order of most other super-resolution techniques^{3,6,21} and slower than the fastest biological STED recordings⁷, by gathering the signal from typically many molecules located at predefined positions, RESOLFT has all the prerequisites for fast imaging. Scanning with arrays of doughnuts or zero-intensity lines (so-called structured illumination^{5,14,18,46}) and detection by a camera will reduce the number of scanning steps required to cover large fields of view and facilitate low-intensity video-rate imaging. The maximum recording speed is determined by the time it takes to establish the disparity of (on–off) states in space, that is, by the switching kinetics, which probably can be improved by further mutagenesis. Note that the switching is not restricted to changes in brightness (on–off) only. Other reversible transitions between disparate states may also prove suitable for RESOLFT imaging, such as states yielding differences in emission wavelengths, lifetime or polarization.

Photoswitching between long-lived states also poses challenges, because in the process the molecule can assume transient (dark) states, such as triplet states, which depend on the molecular micro-environment. In this regard, STED maintains a unique advantage because it entails just basic optical transitions between the ground and the fluorescent state; no atom relocation, spin flip or change in chemical bond is required to switch the fluorescence capability of the molecule — just light. Therefore, switching fluorescence by STED is nearly universal and instantaneous.

The switching stamina of rsEGFP also enabled writing and reading of patterns of both subdiffraction size and spacing d , which has so far been difficult for direct far-field optical writing. In our study, the smallest obtainable structure size was co-determined by the fact that the 532-nm light moderately bleached the off-state proteins too, thus reducing the writing contrast. However, this initial demonstration should spur on new advancements in this field, because current nanowriting efforts are dominated by concepts that resort to much shorter wavelengths of electromagnetic radiation at which focusing becomes exceedingly difficult. In fact, RESOLFT and related concepts are unique for creating materials that are nanostructured in three dimensions³⁰. To maximize the resolution along the optical axis (z), RESOLFT imaging and writing can also be combined with 4Pi microscopy⁴⁷, in which case three-dimensional resolution of $< 10 \text{ nm}$ should become possible at ultralow light levels.

The resolution demonstrated here is similar or even exceeds the resolution attained until now by STED in living cells^{8,10}. Although in both methods the resolution can be continually increased by increasing I_m/I_s , in STED microscopy this strategy will reach practical limits due to the intensities required. Using a threshold intensity I_s that is lower by many orders of magnitude, switching between long-lived states overcomes these limits and, as we have demonstrated here, offers a pathway to lens-based optical imaging and writing at molecular dimensions.

METHODS SUMMARY

Protein generation and screening. Site-directed mutagenesis was performed with the QuikChange Site Directed Mutagenesis Kit (Stratagene) or a multiple-site approach using several degenerative primers. The proteins were expressed from the high-copy expression vector pQE31 (Qiagen) and expressed in *E. coli*. **Viral transfection.** A modified Semliki Forest Virus containing the pSCA-Lifect-rsEGFP vector construct was injected into the slice cultures using a patch pipette. Imaging was performed within 16–48 h after incubation.

Data storage. A layer containing immobilized rsEGFP was prepared by mixing 24.5 μl purified proteins (0.09 mM) with 17.5 μl Tris-HCl pH 7.5, 30 μl acrylamide (Rotiphorese Gel 30, Roth), 0.75 μl 10% ammonium persulfate and 1 μl 10%

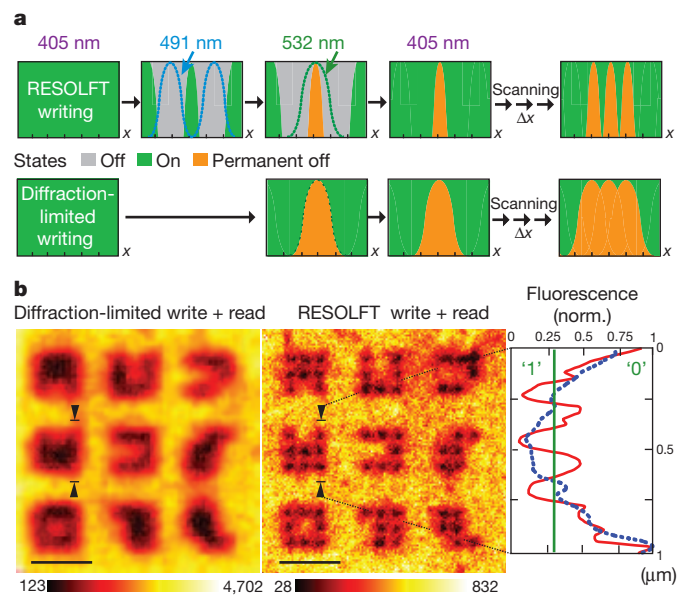


Figure 4 | Subdiffraction-resolution writing and reading using rsEGFP and visible light. **a**, Top, schematic of RESOLFT writing: rsEGFP molecules are switched off at 491 nm using a doughnut-shaped focal intensity (dashed blue line) so that the on-state is confined to a subdiffraction-sized region around the doughnut centre. Subsequent irradiation with 532-nm light makes the on-state molecules permanent by bleaching. Irradiation at 405 nm switches the off-state molecules back into the on-state, allowing the writing of another feature in subdiffraction proximity. Bottom, schematic of diffraction-limited writing. **b**, Conventional (left) and subdiffraction RESOLFT (middle) joined writing and reading in a layer of immobilized rsEGFP. The outlines of the corresponding 3×3 bit patterns were identical. The distance between two bleached spots was 250 nm in each case. Right, normalized line profiles of the fluorescence signal between the two arrows (solid red, RESOLFT; dashed blue, confocal). Scale bars, 1 μm .

TEMED. About 10 µl of this solution was placed on a glass slide and a cover slip was pressed onto the sample to attain a thin layer. Custom MATLAB (The MathWorks) programs allowed automated generation of the voltages and signals for moving the sample and for generating the desired laser pulses. Images were also taken using the software Inspector (<http://www.inspector.de>).

RESOLFT set-up. We implemented a home-built confocal microscope with a normally focused beam for generating fluorescence plus a doughnut-shaped beam for switching rsEGFP off (both at 491 nm wavelength). The beams were circularly polarized, superimposed in the focal plane and applied sequentially. The 405 nm beam for switching rsEGFP on was also circularly polarized. The fluorescence emitted between 500–560 nm was imaged on the opening of a multi-mode fibre and detected by a counting avalanche photodiode. The same set-up was used for writing, which was most specific at 532 nm.

Full details of methods used are available in Supplementary Information.

Received 23 May; accepted 24 August 2011.

Published online 11 September 2011.

- Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).
- Hell, S. W. & Kroug, M. Ground-state depletion fluorescence microscopy, a concept for breaking the diffraction resolution limit. *Appl. Phys. B* **60**, 495–497 (1995).
- Hell, S. W. Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007).
- Klar, T. A., Jakobs, S., Dyba, M., Egner, A. & Hell, S. W. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proc. Natl Acad. Sci. USA* **97**, 8206–8210 (2000).
- Hell, S. W. Toward fluorescence nanoscopy. *Nature Biotechnol.* **21**, 1347–1355 (2003).
- Hell, S. W. Microscopy and its focal switch. *Nature Methods* **6**, 24–32 (2009).
- Westphal, V. *et al.* Video-rate far-field optical nanoscopy dissects synaptic vesicle movement. *Science* **320**, 246–249 (2008).
- Hein, B., Willig, K. I. & Hell, S. W. Stimulated emission depletion (STED) nanoscopy of a fluorescent protein-labeled organelle inside a living cell. *Proc. Natl Acad. Sci. USA* **105**, 14271–14276 (2008).
- Egeling, C. *et al.* Direct observation of the nanoscale dynamics of membrane lipids in a living cell. *Nature* **457**, 1159–1162 (2009).
- Nägerl, U. V., Willig, K. I., Hein, B., Hell, S. W. & Bonhoeffer, T. Live-cell imaging of dendritic spines by STED microscopy. *Proc. Natl Acad. Sci. USA* **105**, 18982–18987 (2008).
- Hell, S. W., Jakobs, S. & Kastrup, L. Imaging and writing at the nanoscale with focused visible light through saturable optical transitions. *Appl. Phys., A Mater. Sci. Process.* **77**, 859–860 (2003).
- Hell, S. W., Dyba, M. & Jakobs, S. Concepts for nanoscale resolution in fluorescence microscopy. *Curr. Opin. Neurobiol.* **14**, 599–609 (2004).
- Hell, S. W. Strategy for far-field optical imaging and writing without diffraction limit. *Phys. Lett.* **326**, 140–145 (2004).
- Gustafsson, M. G. L. Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. *Proc. Natl Acad. Sci. USA* **102**, 13081–13086 (2005).
- Hell, S. W. in *Topics in Fluorescence Spectroscopy* Vol. 5 (ed. Lakowicz, J. R.) 361–422 (Plenum, 1997).
- Hofmann, M., Egeling, C., Jakobs, S. & Hell, S. W. Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins. *Proc. Natl Acad. Sci. USA* **102**, 17565–17569 (2005).
- Lukyanov, K. A. *et al.* Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog. *J. Biol. Chem.* **275**, 25879–25882 (2000).
- Schwentker, M. A. *et al.* Wide-field subdiffraction RESOLFT microscopy using fluorescent protein photoswitching. *Microsc. Res. Tech.* **70**, 269–280 (2007).
- Ando, R., Mizuno, H. & Miyawaki, A. Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science* **306**, 1370–1373 (2004).
- Dedecker, P. *et al.* Subdiffraction imaging through the selective donut-mode depletion of thermally stable photoswitchable fluorophores: numerical analysis and application to the fluorescent protein dronpa. *J. Am. Chem. Soc.* **129**, 16132–16141 (2007).
- Huang, B., Babcock, H. & Zhuang, X. Breaking the diffraction barrier: super-resolution imaging of cells. *Cell* **143**, 1047–1058 (2010).
- Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
- Hess, S. T., Girirajan, T. P. K. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258–4272 (2006).
- Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3**, 793–796 (2006).
- Dickson, R. M., Cubitt, A. B., Tsien, R. Y. & Moerner, W. E. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature* **388**, 355–358 (1997).
- Bossi, M., Foelling, J., Dyba, M., Westphal, V. & Hell, S. W. Breaking the diffraction resolution barrier in far-field microscopy by molecular optical bistability. *N. J. Phys.* **8**, 275 (2006).
- Scott, T. F., Kowalski, B. A., Sullivan, A. C., Bowman, C. N. & McLeod, R. R. Two-color single-photon photoinitiation and photoinhibition for subdiffraction photolithography. *Science* **324**, 913–917 (2009).
- Li, L., Gattass, R. R., Gershgoren, E., Hwang, H. & Fourkas, J. T. Achieving 1/20 resolution by one-color initiation and deactivation of polymerization. *Science* **324**, 910–913 (2009).
- Andrew, T. L., Tsai, H. Y. & Menon, R. Confining light to deep subwavelength dimensions to enable optical nanopatterning. *Science* **324**, 917–921 (2009).
- Fischer, J., Freymann, G. & Wegener, M. The materials challenge in diffraction-unlimited direct-laser-writing optical lithography. *Adv. Mater.* **22**, 3578–3582 (2010).
- Ormö, M. *et al.* Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392–1395 (1996).
- Andresen, M. *et al.* Structure and mechanism of the reversible photoswitch of a fluorescent protein. *Proc. Natl Acad. Sci. USA* **102**, 13070–13074 (2005).
- Andresen, M. *et al.* Structural basis for reversible photoswitching in Dronpa. *Proc. Natl Acad. Sci. USA* **104**, 13005–13009 (2007).
- Henderson, J. N., Ai, H. W., Campbell, R. E. & Remington, S. J. Structural basis for reversible photobleaching of a green fluorescent protein homologue. *Proc. Natl Acad. Sci. USA* **104**, 6672–6677 (2007).
- Adam, V. *et al.* Structural characterization of IrisFP, an optical highlighter undergoing multiple photo-induced transformations. *Proc. Natl Acad. Sci. USA* **105**, 18343–18348 (2008).
- Brakemann, T. *et al.* Molecular basis of the light-driven switching of the photochromic fluorescent protein Padron. *J. Biol. Chem.* **285**, 14603–14609 (2010).
- Patterson, G. H., Knobel, S. M., Sharif, W. D., Kain, S. R. & Piston, D. W. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys. J.* **73**, 2782–2790 (1997).
- Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913–916 (2002).
- Bizzarri, R. *et al.* Single amino acid replacement makes *Aequorea victoria* fluorescent proteins reversibly photoswitchable. *J. Am. Chem. Soc.* **132**, 85–95 (2010).
- Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
- Adam, V. *et al.* Data storage based on photochromic and photoconvertible fluorescent proteins. *J. Biotechnol.* **149**, 289–298 (2010).
- Vats, P. & Rothfield, L. Duplication and segregation of the actin (MreB) cytoskeleton during the prokaryotic cell cycle. *Proc. Natl Acad. Sci. USA* **104**, 17795–17800 (2007).
- Riedl, J. *et al.* Lifeact: a versatile marker to visualize F-actin. *Nature Methods* **5**, 605–607 (2008).
- Harris, K. M. & Kater, S. B. Dendritic spines: cellular specializations imparting both stability and flexibility to synaptic function. *Annu. Rev. Neurosci.* **17**, 341–371 (1994).
- Hopt, A. & Neher, E. Highly nonlinear photodamage in two-photon fluorescence microscopy. *Biophys. J.* **80**, 2029–2036 (2001).
- Heintzmann, R., Jovin, T. M. & Cremer, C. Saturated patterned excitation microscopy—a concept for optical resolution improvement. *JOSA A* **19**, 1599–1609 (2002).
- Hell, S. W., Schmidt, R. & Egner, A. Diffraction-unlimited three-dimensional optical nanoscopy with opposing lenses. *Nature Photon.* **3**, 381–387 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Jethwa for careful reading and M. Andresen, T. Brakemann, S. Löbermann, R. Schmitz-Salue and A. C. Stiel for discussions and support, as well as T. Gilat and F. Voss (MPI of Neurobiology, Munich) for help with the slice culture preparation and A. Schönke for adapting the software Inspector. We thank The Project Gutenberg for making Grimm's Fairy Tales available in electronic format, L. Rothfield (University of Connecticut Health Center) for providing the plasmid pLE7, R. Wedlich-Soldner (MPI of Biochemistry, Munich) for the lifeact-YFP construct and V. Stein (MPI of Neurobiology, Munich) for the virus protocol. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) through the DFG-Research Center for Molecular Physiology of the Brain (to S.J.) and by a Gottfried-Wilhelm-Leibniz prize of the DFG (to S.W.H.).

Author Contributions T.G., I.T., M.L., H.B., F.L.-C. performed research, I.T., M.L., T.G., H.B., C.E. set up the microscopes, N.T.U., K.I.W. prepared samples, M.L., T.G., I.T., K.I.W., S.J., S.W.H. analysed data, S.J., C.E., S.W.H. designed research. S.J., M.L., S.W.H. wrote the paper. All authors discussed the data and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.J. (sjakobs@gwdg.de) or S.W.H. (shell@gwdg.de).

The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus

Yue Zhao^{1,2*}, Jieliang Yang^{1,2*}, Jianjin Shi², Yi-Nan Gong², Qiuhe Lu², Hao Xu², Liping Liu² & Feng Shao²

Inflammasomes are large cytoplasmic complexes that sense microbial infections/danger molecules and induce caspase-1 activation-dependent cytokine production and macrophage inflammatory death^{1,2}. The inflammasome assembled by the NOD-like receptor (NLR) protein NLRC4 responds to bacterial flagellin and a conserved type III secretion system (TTSS) rod component^{3–5}. How the NLRC4 inflammasome detects the two bacterial products and the molecular mechanism of NLRC4 inflammasome activation are not understood. Here we show that NAIP5, a BIR-domain NLR protein required for *Legionella pneumophila* replication in mouse macrophages⁶, is a universal component of the flagellin–NLRC4 pathway. NAIP5 directly and specifically interacted with flagellin, which determined the inflammasome-stimulation activities of different bacterial flagellins. NAIP5 engagement by flagellin promoted a physical NAIP5–NLRC4 association, rendering full reconstitution of a flagellin-responsive NLRC4 inflammasome in non-macrophage cells. The related NAIP2 functioned analogously to NAIP5, serving as a specific inflammasome receptor for TTSS rod proteins such as *Salmonella* PrgJ and *Burkholderia* BsaK. Genetic analysis of *Chromobacterium violaceum* infection revealed that the TTSS needle protein CprI can stimulate NLRC4 inflammasome activation in human macrophages. Similarly, CprI is specifically recognized by human NAIP, the sole NAIP family member in human. The finding that NAIP proteins are inflammasome receptors for bacterial flagellin and TTSS apparatus components further predicts that the remaining NAIP family members may recognize other unidentified microbial products to activate NLRC4 inflammasome-mediated innate immunity.

The NLR protein NLRC4 (also known as IPAF) in macrophages activates caspase-1 and downstream inflammatory response upon sensing cytosolic presence of flagellin during bacterial infection^{3,4,7,8}. To study the mechanism of the NLRC4 inflammasome, a defined biochemical assay was developed by fusing recombinant flagellin carboxy-terminal to the amino-terminal domain of anthrax lethal factor. This domain (designated as LFn here), through binding to another anthrax protein called protective antigen (PA), can efficiently translocate heterologous fusion proteins into mammalian cytosol through endocytosis-mediated entry⁹. Using this system, purified flagellin from *L. pneumophila* (LFn-FlaA^{LP}) was found to trigger robust caspase-1 cleavage (Fig. 1a), IL-1 β release and pyroptotic death (Supplementary Fig. 1a, b) in primary bone-marrow-derived macrophages (BMMs). These activations were completely diminished in *Nlrp4*^{−/−} and caspase-1^{−/−} macrophages (Fig. 1a and Supplementary Fig. 1a, b); *Asc*^{−/−} (*Asc* also known as *Pycard*) BMMs also showed little caspase-1 maturation and IL-1 β release but with a partially affected pyroptosis due to ASC-independent NLRC4 inflammasome activation¹⁰. Full activation of NLRC4 inflammasome requires L470, L472 and L473 in *Legionella* flagellin⁶. Accordingly, alanine substitutions of the three leucine residues (3A) generated a largely inactive LFn-FlaA^{LP} protein (Fig. 1a and Supplementary Fig. 1a, b). LFn-FlaA^{LP} induced similar NLRC4-dependent caspase-1 activation and pyroptosis in

Tlr4^{−/−} macrophages (Supplementary Figs 2a, b and 3a), excluding a possible contribution from residual endotoxin contaminants present in the recombinant protein. Other bacteria such as *Salmonella typhimurium* also trigger flagellin-dependent NLRC4 inflammasome activation^{3,4,8}. Delivery of *S. typhimurium* (LFn-FliCSt) or *Yersinia enterocolitica* (LFn-FliC2^{Ye}) flagellin into BMMs induced robust caspase-1 activation and extensive pyroptosis in an NLRC4-dependent manner (Supplementary Fig. 4a, b). Thus, LFn-mediated delivery of recombinant flagellin recapitulates all known genetic properties of flagellin activation of the NLRC4 inflammasome.

For *L. pneumophila* infection, flagellin-induced caspase-1 activation requires NAIP5 (also known as BIRC1E), a BIR-domain-containing NLR protein⁶. A natural variant of NAIP5 renders macrophages from the A/J mouse permissive to *L. pneumophila* intracellular replication^{11–16}. The role of NAIP5 for other bacterial flagellins is not clear^{6,17}. RNA interference (RNAi) knockdown of *Naip5* (Supplementary Fig. 3b) severely blocked LFn-FlaA^{LP}-triggered caspase-1 activation and pyroptosis (Supplementary Fig. 2a, b). Notably, activation of the NLRC4 inflammasome by LFn-FliCSt and LFn-FliC2^{Ye}, but not that by *Salmonella* TTSS rod protein (LFn-PrgJ), was also drastically reduced by short hairpin RNA (shRNA)-mediated stable knockdown of *Naip5* (Fig. 1b, c and Supplementary Fig. 3c). Consistently, flagellin-triggered caspase-1 activation during *Salmonella* and *Legionella* infection was significantly attenuated in *Naip5* knockdown macrophages (Fig. 1d).

The finding that NAIP5 is a possible integral component of the flagellin–NLRC4 pathway inspired us to investigate whether NAIP5 directly recognizes flagellin. *Legionella* flagellin was found to show an evident yeast two-hybrid interaction with NAIP5, but not NLRC4, whereas the 3A mutant showed no interaction (Fig. 2a). *Naip5* is located within a genomic locus containing seven highly homologous *Naip* genes (*Naip1–7*) and four of them (*Naip1*, *Naip2*, *Naip5* and *Naip6*) have transcripts in C57BL/6 mice¹¹. *Legionella* flagellin also showed a two-hybrid interaction with NAIP6, but not with NAIP1 and NAIP2 (Fig. 2a). Supporting the two-hybrid results, *Legionella* flagellin expressed in 293T cells readily co-precipitated NAIP5 and NAIP6, but not NAIP1, NAIP2 and NLRC4, whereas the 3A mutant failed to do so (Fig. 2b). The TLR5-binding-deficient mutant (I391A), which is fully functional in inflammasome activation⁶, behaved similarly to wild-type flagellin in the co-immunoprecipitation assay. Flagellin also co-precipitated NAIP5 encoded by the A/J allele (NAIP5^{A/J}) (Fig. 2b), which explains the normal or nearly normal caspase-1 activation in *L. pneumophila*-infected A/J macrophages^{15,18,19}.

A panel of nine additional flagellins from different bacteria was further profiled (Fig. 2c). In the two-hybrid assay, flagellins from *S. typhimurium*, *Y. enterocolitica*, *Photobacterium luminescens* and *Pseudomonas aeruginosa* showed a positive result whereas those from enteropathogenic *Escherichia coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), *Shigella flexneri*, *Chromobacterium violaceum* and *Burkholderia thailandensis* did not interact with NAIP5 (Fig. 2c and Supplementary Fig. 5a). NAIP5 interaction with *S. typhimurium* flagellin

¹Graduate Program in Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China. ²National Institute of Biological Sciences, Beijing, 102206, China.

*These authors contributed equally to this work.

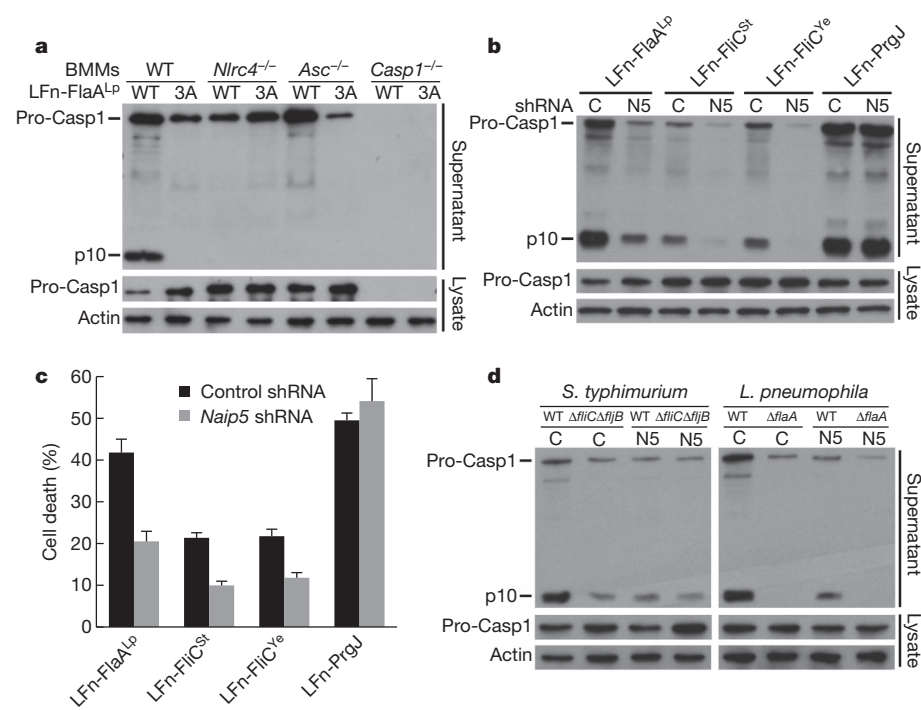


Figure 1 | A defined biochemical assay reveals a universal role of NAIP5 in flagellin-triggered NLRC4 inflammasome activation in mouse macrophages. **a**, Effects of anthrax lethal factor N-terminal-domain-mediated intracellular delivery of *Legionella* flagellin (LFn-FlaA^{Lp}) on caspase-1 activation in lipopolysaccharide (LPS)-primed BMMs derived from wild-type (WT, C57BL/6) or indicated knockout mice. 3A denotes a triple mutant flagellin (L470A/L472A/L473A). Shown are anti-caspase-1 and anti-actin immunoblots of culture supernatants (top) and total cell lysates (bottom). p10 denotes the processed mature form of caspase-1. **b**, **c**, Effects of *Naip5* knockdown on flagellin-induced caspase-1 activation (**b**) and cell death (**c**). A

Naip5-targeting (N5) (Supplementary Table 1) or a control (C) shRNA was stably expressed in immortalized BMMs. LFn-FlaA^{Lp}, FliCSt and FliC^{Ye} are recombinant LFn-tagged flagellins from *L. pneumophila*, *S. typhimurium* and *Y. enterocolitica*, respectively. LFn-PrgJ is LFn-tagged *S. typhimurium* TTSS rod protein. **c**, LDH releases are shown as mean values \pm standard deviation (s.d.) from three independent determinations. **d**, Effects of *Naip5* knockdown on flagellin-induced caspase-1 activation during *Salmonella* and *Legionella* infection. Δ fliC Δ fliB and Δ fliA denote flagellin-deficient strains of *S. typhimurium* and *L. pneumophila*, respectively.

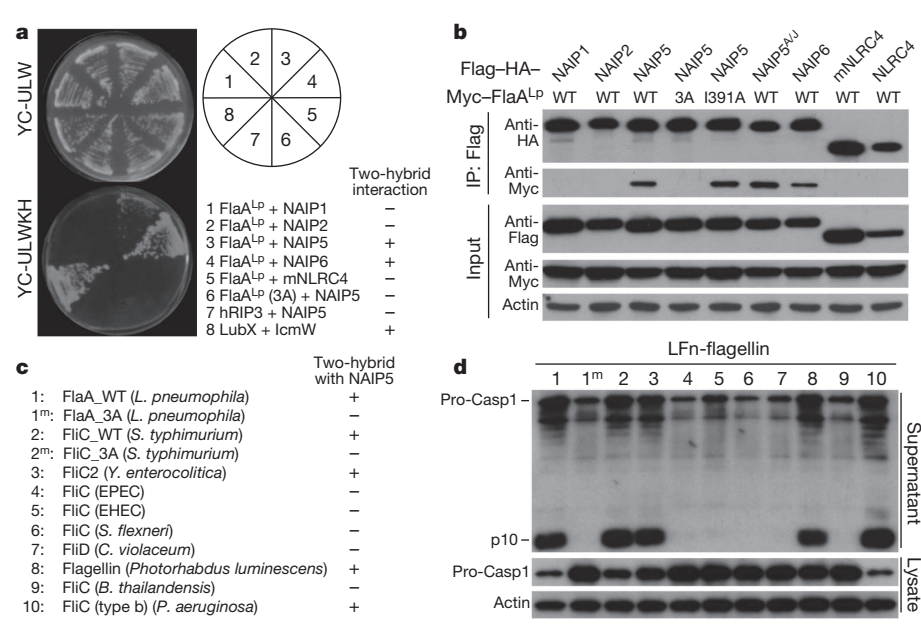


Figure 2 | Flagellin interacts specifically with NAIP5 and the interaction correlates with the activity of flagellins from different bacteria. **a**, Yeast two-hybrid interaction assays of *Legionella* flagellin (FlaA^{Lp}) and different NAIP proteins (or mouse (m)NLRC4). The chart in the lower right corner summarizes the interaction results. The known interaction between *Legionella* effector LubX and its secretion chaperon IcmW was included as a positive control. **b**, Co-immunoprecipitation assays of *Legionella* flagellin (FlaA^{Lp}) and

different NAIP proteins (or NLRC4). Shown are immunoblots of anti-Flag immunoprecipitates (IP: Flag) and total cell lysates (Input). I391A is a TLR5 binding-deficient flagellin mutant. **c**, Summary of yeast two-hybrid interaction of NAIP5 with different bacterial flagellins. The raw data are shown in Supplementary Fig. 5a. **d**, Caspase-1 activation assays of LFn-mediated delivery of different bacterial flagellins into primary BMMs. Number denotations of different flagellins follow those in **c**.

required three leucine residues equivalent to those in *Legionella* flagellin. When delivered into BMMs, flagellins from *S. typhimurium*, *Y. enterocolitica*, *P. luminescens* and *P. aeruginosa*, but not those from the other five bacteria species, stimulated caspase-1 activation, macrophage death and IL-1 β release (Fig. 2d and Supplementary Fig. 5b, c). The positive NAIP5-binding and inflammasome-stimulating activities of *S. typhimurium* and *P. aeruginosa* flagellins agree with their genetic requirements for infection-induced caspase-1 activation^{3,4,8,20–22}. Among those inactive ones, *S. flexneri* flagellin is not expressed and dispensable for host innate immune detection of *S. flexneri* infection²³. Genetic ablations of flagellins from EPEC and *B. thailandensis* also did not affect infection-induced caspase-1 activation (Supplementary Fig. 6). Thus, the ability of the ten different flagellins to interact with NAIP5 correlates well with their differential inflammasome-stimulating activity, which further supports the idea that flagellin is generally recognized by NAIP5 in triggering NLRC4 inflammasome activation.

NAIP5 and NLRC4 were then co-expressed in 293T cells and their possible interactions were investigated. Co-immunoprecipitation of NAIP5 and NLRC4 was barely detectable in the absence of flagellin. However, co-expression of *Legionella* flagellin, but not the 3A mutant, significantly increased the amount of NAIP5 precipitated by NLRC4 (Fig. 3a, b). Flagellin was also detected in the NLRC4 immunoprecipitates due to the bridging effect of NAIP5. Deletion of the nucleotide-binding P-loop in NLRC4 nucleotide-binding and oligomerization domain (NOD) abolished flagellin-stimulated NLRC4–NAIP5 interaction (Fig. 3a), which agrees with the reported interaction between NOD domains from NLRC4 and NAIP5 (ref. 24). Flagellin also promoted the association of NLRC4 with NAIP5^{A/J}, but neither NAIP1 nor NAIP2 was precipitated by NLRC4 despite the presence of flagellin (Fig. 3b). To test whether the flagellin-stimulated NAIP5–NLRC4 complex can activate downstream signalling, NAIP5 and NLRC4, together with pro-caspase-1 and pro-IL-1 β , were co-expressed in 293T cells. Delivery of LFn-FlaA^{LP}, but not the 3A mutant, into the transfected cells resulted in an evident production of mature IL-1 β (Fig. 3c and Supplementary Fig. 7a). Omission of NAIP5, NLRC4 or caspase-1 in this reconstitution abolished the response to flagellin

stimulation. NAIP5^{A/J} also supported the reconstitution whereas NAIP1 and NAIP2 failed to do so (Fig. 3d and Supplementary Fig. 7b), consistent with their differential association with NLRC4 upon flagellin stimulation (Fig. 3b). Moreover, the reconstituted NLRC4 inflammasome exhibited robust responses to flagellins from *Salmonella* and *Yersinia* (Supplementary Fig. 8a). Each of the three domains in both NLR proteins (CARD, NOD and LRR in NLRC4; BIR, NOD and LRR in NAIP5) was essential for assembling a flagellin-responsive inflammasome complex (Supplementary Fig. 8b). These results indicate that flagellin recognition by NAIP5 stimulates the physical association between NAIP5 and NLRC4, thereby signalling downstream caspase-1 activation.

NAIP6 interacted with flagellin in a manner similar to NAIP5 (Fig. 2b and Supplementary Fig. 9) and supported the reconstitution in 293T cells (Fig. 3b, d). In fact, NAIP6, among all the NAIP proteins, shares the highest sequence identity with NAIP5, of 94.7% (Supplementary Fig. 10). NAIP6 probably has a similar function to NAIP5 in macrophage detection of flagellin, but its role might be relatively minor given its much lower expression in primary macrophages compared with that of NAIP5 (ref. 12).

The NLRC4 inflammasome also responds to a conserved TTSS rod protein such as PrgJ in *S. typhimurium*, BsaK in *B. thailandensis* and EscI in EPEC⁵. Delivery of recombinant BsaK (LFn-BsaK) into BMMs recapitulated such effects and induced NLRC4-dependent caspase-1 activation and pyroptosis (Supplementary Fig. 11). Given that NAIP5 recognizes flagellin and that PrgJ activation of the NLRC4 inflammasome is independent of NAIP5 (ref. 17), we proposed that other NAIP proteins could recognize the TTSS rod protein. BsaK was found to interact with NAIP2, but not NAIP1, NAIP5, NAIP6 and NLRC4, in the two-hybrid assay (Fig. 4a). Co-immunoprecipitation assay confirmed this NAIP2-specific interaction (Fig. 4b). This observation agrees with the idea that NAIP2 is the most distantly related to the other NAIP proteins (Supplementary Fig. 10). Reconstitution in non-macrophage cells further showed that only NAIP2, but not any other NAIP, effected robust IL-1 β maturation upon LFn-BsaK stimulation (Fig. 4c). These findings indicate that NAIP2 is the specific receptor for the TTSS rod protein.

To test the requirement of NAIP2 for detecting the TTSS rod protein in macrophages, *Naip2* stable knockdown BMMs were generated. Among the four different shRNAs (*Naip2-1*, 2, 3 and 4), *Naip2-1* and *Naip2-2* considerably reduced *Naip2* messenger RNA level whereas *Naip2-3* and *Naip2-4* showed intermediate and negligible efficiency, respectively (Supplementary Fig. 12a). *Naip2-1* and *Naip2-2* knockdown macrophages exhibited significant resistance in caspase-1 activation and pyroptosis to LFn-PrgJ or LFn-BsaK stimulation (Supplementary Fig. 12b–d). In contrast, *Naip2-3* knockdown macrophages showed a mild resistance and *Naip2-4* knockdown macrophages had a normal sensitivity to rod protein stimulations. In *Naip2-2* knockdown macrophages, in which mRNA levels of other *Naip* genes were not affected (Supplementary Fig. 13), attenuated caspase-1 activation was only observed with the rod protein stimulations, but not with flagellin stimulations (Fig. 4d). Furthermore, deletion of genes encoding the rod proteins from flagellin-deficient EPEC and *S. typhimurium* abolished bacterial infection-induced caspase-1 activation, and this effect did not occur in *Naip2-2* knockdown macrophages (Fig. 4e). These results demonstrate the critical and specific role of NAIP2 in recognizing the TTSS rod protein for NLRC4 inflammasome activation.

In contrast to mouse macrophages, human U937 monocyte-derived macrophages were unresponsive to intracellular delivery of flagellin and BsaK/PrgJ-like rod protein (Supplementary Fig. 14). When profiling our genetic collection of various pathogenic bacteria, a *C. violaceum* strain (deficient in secretion of TTSS effectors) was identified to be capable of inducing caspase-1 activation in human U937 monocytes (Supplementary Fig. 15a). Notably, further ablation of five possible flagellin genes (ΔF) caused no reduction in this activation. Stable knockdown of *NLRC4* (Supplementary Fig. 16) significantly attenuated *C.*

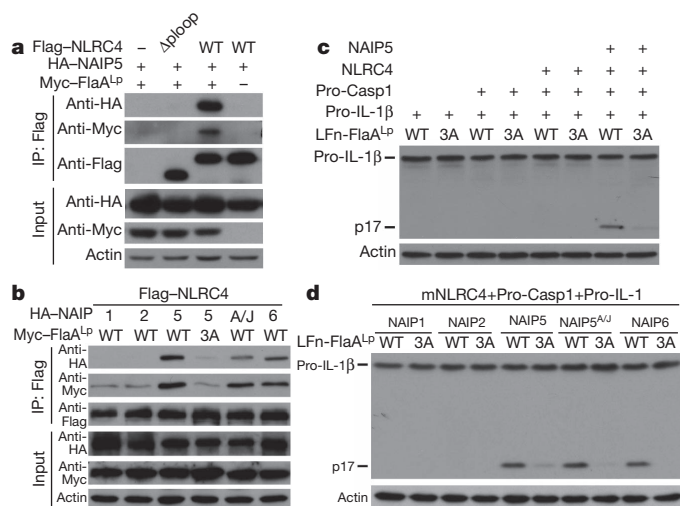


Figure 3 | Flagellin stimulates the NAIP5–NLRC4 association and reconstitution of flagellin activation of the NLRC4 inflammasome in non-macrophage cells. **a, b**, Co-immunoprecipitation assays of NAIP5 and NLRC4 interaction in the presence or absence of flagellin. Δ ploop in **a** denotes an NLRC4 mutant with deletion of the nucleotide-binding P-loop. **c**, Reconstitution of flagellin activation of the NLRC4 inflammasome in non-macrophage cells. Lysates from 293T cells transfected with indicated plasmid combinations and stimulated with LFn-FlaA^{LP} were analysed for mature IL-1 β (p17) by immunoblotting. Expression of transfected inflammasome components for **c** and **d** is in Supplementary Fig. 7. **d**, Assay of different NAIP proteins in supporting reconstitution of flagellin activation of the NLRC4 inflammasome in 293T cells.

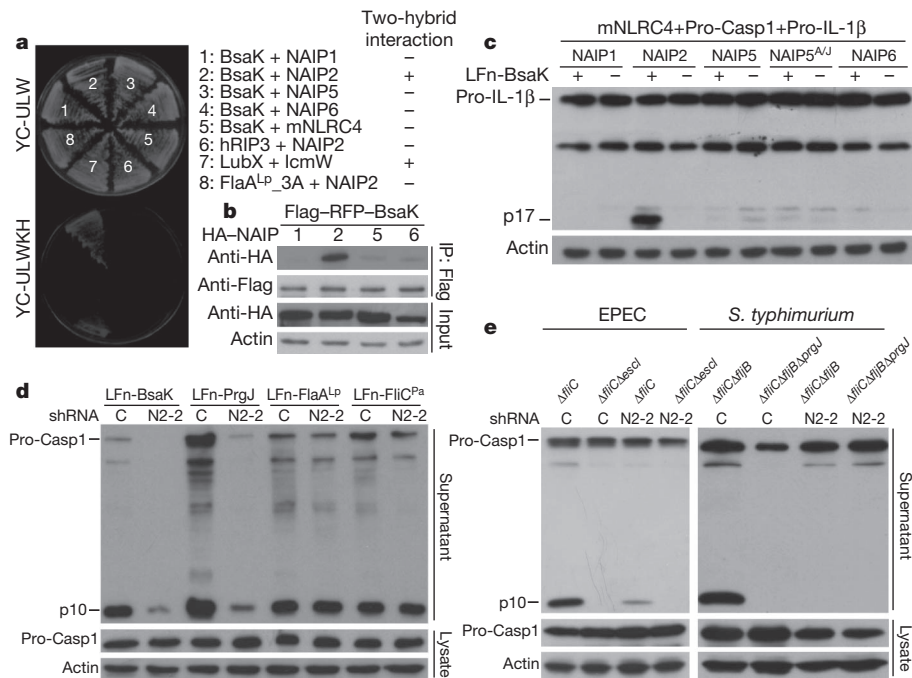


Figure 4 | NAIP2 interacts with the TTSS rod protein and is required for the rod protein to trigger mouse NLRC4 inflammasome activation. **a**, **b**, Yeast two-hybrid (**a**) and co-immunoprecipitation (**b**) assays of interactions between *B. thailandensis* rod protein BsaK and different NAIP proteins.

c, Reconstitution of BsaK activation of the NLRC4 inflammasome in non-macrophage cells. Lysates from HeLa cells transfected with indicated plasmid combinations and stimulated with LFn-BsaK were analysed for mature IL-1 β (p17) by immunoblotting. Expression of transfected inflammasome components is in Supplementary Fig. 7. **d**, Effects of *Naip2* knockdown on

violaceum ΔF -triggered caspase-1 activation and pyroptosis (Fig. 5a and Supplementary Fig. 17a). The PrgJ homologue in the *C. violaceum* Cpi-1 TTSS system, CprI, is encoded in a separate Cpi-1a locus that harbours several additional TTSS apparatus genes²⁵ (Fig. 5b). Although

caspase-1 activation induced by TTSS rod proteins and flagellins. Control (C) or *Naip2*-2 (N2-2) stable knockdown macrophages (Supplementary Fig. 12) were stimulated with purified LFn-tagged BsaK, PrgJ, FlaA^{LP} or FliC^{Pa} proteins as indicated. **e**, Effects of *Naip2* knockdown on rod-protein-induced caspase-1 activation during EPEC and *Salmonella* infection. EPEC E2348/69 $\Delta fliC\Delta escI$ and *S. typhimurium* $\Delta fliC\Delta fliB\Delta prgJ$ denote the rod-protein-deficient EPEC and *S. typhimurium* strains, respectively, which were constructed on the flagellin-deletion background.

cprI was not required for infection-induced caspase-1 activation and pyroptosis, deletion of the entire Cpi-1a locus largely diminished *C. violaceum*-induced NLRC4 inflammasome activation (Fig. 5a, Supplementary Fig. 15b and Supplementary Fig. 17). Further genetic analysis of the entire Cpi-1a locus identified *cprI*, which was essential for inducing caspase-1 activation and pyroptosis (Fig. 5c and Supplementary Fig. 17b). A CprI-expressing plasmid could rescue the deficiencies of inflammasome activation for both *cprI* and Cpi-1a deletion strains (Fig. 5d). Thus, *C. violaceum* requires *cprI* to stimulate NLRC4 inflammasome activation in human macrophages.

cprI encodes the conserved TTSS needle subunit that is a sequence paralogue of the rod protein²⁶, raising a hypothesis that the needle protein is the bacterial ligand recognized by the human NLRC4 inflammasome. Consistent with the above genetic analyses, LFn-mediated delivery of CprI, but not other Cpi-1a-encoded proteins, triggered robust caspase-1 activation and pyroptosis in U937 cells

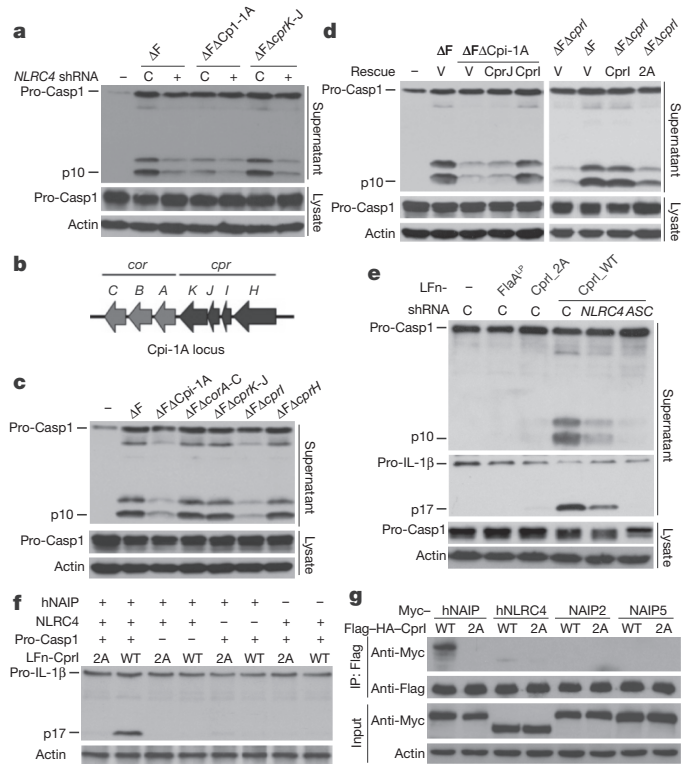


Figure 5 | *C. violaceum* infection studies reveal that the human NLRC4 inflammasome responds to the TTSS needle subunit through specific recognition by human NAIP. **a**–**c**, Caspase-1 activation assays of *C. violaceum* infections of human U937 monocyte-derive macrophages. ΔF has deletions of five possible flagellin genes in *C. violaceum*. Control (C) or NLRC4 (+) stable knockdown cells were used in **a**. $\Delta F\Delta Cpi-1A$ means a deletion of the entire TTSS Cpi-1A locus illustrated in the schematic drawing shown in **b**. Detailed information for all the mutant strains are listed in Supplementary Table 3.

d, Complementation of Cpi-1a locus or *cprI* deletion *C. violaceum* strains by a CprI-expressing plasmid. PMA-differentiated U937 cells were infected with indicated *C. violaceum* mutant or rescue strain. 2A is a double mutant of CprI (V69A/I79A). **e**, Caspase-1 activation assays of delivery of CprI into human U937 macrophages and effects of NLRC4 and ASC knockdown. Control (C) or NLRC4 or ASC stable knockdown cells were stimulated with LFn-CprI or other indicated LFn fusion proteins. **f**, Reconstitution of CprI activation of the human NLRC4 inflammasome in 293T cells. **g**, Co-immunoprecipitation assay of CprI and different NAIP proteins (or NLRC4).

(Supplementary Fig. 18a, b), which were largely decreased in *NLRC4* and ASC knockdown cells (Fig. 5e and Supplementary Fig. 18c). Mutation of two hydrophobic residues (V69A/I79A, 2A) in a helical hairpin region in CprI diminished its activity of stimulating inflammasome activation (Fig. 5d, e and Supplementary Fig. 18). CprI activation of the NLRC4 inflammasome could also be robustly reconstituted in 293T cells and the 2A mutant remained inactive in this assay (Fig. 5f). Most importantly, this reconstitution required human NAIP, the sole NAIP family member in human. Human NAIP-based reconstitution specifically responded to LFn-CprI, but not to LFn-FlaA^{LP} and LFn-BsaK; LFn-CprI did not activate NAIP5- and NAIP2-based reconstitution (Supplementary Fig. 19). Furthermore, CprI readily co-precipitated human NAIP, but not any of NAIP2, NAIP5 and NLRC4 from 293T cells, and the nonfunctional 2A mutant failed to interact with human NAIP (Fig. 5g). Homologous needle subunits from EHEC, *B. thailandensis*, *P. aeruginosa*, *S. flexneri* and *S. typhimurium*, but not those from EPEC and *V. parahaemolyticus*, also stimulated NLRC4 inflammasome activation in U937 cells (Supplementary Fig. 20). Thus, human NAIP functions analogously to mouse NAIP5/2, but specifically recognizes the TTSS needle subunit to trigger human NLRC4 inflammasome activation.

In summary, murine NLR proteins NAIP5 and NAIP2 directly recognize bacterial flagellin and TTSS rod protein, respectively, whereas human NAIP serves as a specific receptor for the TTSS needle protein. Engagement of NAIP receptors by corresponding bacterial ligands promotes their physical association with NLRC4, resulting in activation of the NLRC4 inflammasome and macrophage innate immunity. The inflammasome-stimulating activities of flagellin, TTSS rod and needle proteins lie in their C-terminal leucine-rich helical hairpin regions that share structural commonalities^{5,27}. Thus, other homologous NAIP proteins might recognize additional bacterial products of similar biochemical features for counteracting diverse bacterial infections. Our results also indicate that NLRC4 acts as an adaptor through which inflammasome activation signals generated from different NAIP receptors are transduced to caspase-1. Involvement of an additional cytosolic pattern recognition receptor (PRR) protein for sensing one microbial product has previously been noted with NLRP3 and NALP1-mediated inflammasome activation^{28,29}. Future studies will probably identify more PRR proteins that act sequentially within a single inflammasome complex in response to microbial products or danger signals.

METHODS SUMMARY

LFn-mediated intracellular delivery and RNAi. For delivery into macrophages, purified recombinant proteins were washed with 60% isopropanol to remove the majority of endotoxin contaminants. LFn-flagellin, LFn-BsaK/PrgI/CprI, LFn-CprI or other indicated control proteins together with PA proteins were added into culture medium (serum-free) at a final concentration of 1 µg ml⁻¹ for each protein. Cells were further incubated for 1 h (primary BMMs) or 3 h (immortalized BMMs) before being subjected to the indicated inflammasome activation assays. Transient small interfering RNA (siRNA) knockdown in macrophages was performed using the INTERFERin reagent (Polyplus Transfection) by following the manufacturer's instruction. To achieve stable knockdown in macrophages, a modified pLKO.1-GFP plasmid harbouring a specific shRNA (Supplementary Table 1) was transduced into BMMs or U937 cells by lentiviral infection and GFP-positive knockdown cells were sorted out by flow cytometry for further functional analysis.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 May; accepted 30 August 2011.

Published online 14 September 2011.

1. Lamkanfi, M. & Dixit, V. M. Inflammasomes: guardians of cytosolic sanctity. *Immunity* **27**, 95–105 (2009).
2. Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821–832 (2010).
3. Franchi, L. et al. Cytosolic flagellin requires Ipaf for activation of caspase-1 and interleukin 1β in salmonella-infected macrophages. *Nature Immunol.* **7**, 576–582 (2006).
4. Miao, E. A. et al. Cytoplasmic flagellin activates caspase-1 and secretion of interleukin 1β via Ipaf. *Nature Immunol.* **7**, 569–575 (2006).

5. Miao, E. A. et al. Innate immune detection of the type III secretion apparatus through the NLRC4 inflammasome. *Proc. Natl Acad. Sci. USA* **107**, 3076–3080 (2010).
6. Lightfield, K. L. et al. Critical function for Naip5 in inflammasome activation by a conserved carboxy-terminal domain of flagellin. *Nature Immunol.* **9**, 1171–1178 (2008).
7. Amer, A. et al. Regulation of Legionella phagosome maturation and infection through flagellin and host Ipaf. *J. Biol. Chem.* **281**, 35217–35223 (2006).
8. Broz, P. et al. Redundant roles for inflammasome receptors NLRP3 and NLRC4 in host defense against *Salmonella*. *J. Exp. Med.* **207**, 1745–1755 (2010).
9. Milne, J. C., Blanke, S. R., Hanna, P. C. & Collier, R. J. Protective antigen-binding domain of anthrax lethal factor mediates translocation of a heterologous protein fused to its amino- or carboxy-terminus. *Mol. Microbiol.* **15**, 661–666 (1995).
10. Broz, P., von Moltke, J., Jones, J. W., Vance, R. E. & Monack, D. M. Differential requirement for Caspase-1 autoproteolysis in pathogen-induced cell death and cytokine processing. *Cell Host Microbe* **8**, 471–483 (2010).
11. Diez, E. et al. *Birc1e* is the gene within the *Lgn1* locus associated with resistance to *Legionella pneumophila*. *Nature Genet.* **33**, 55–60 (2003).
12. Wright, E. K. et al. Naip5 affects host susceptibility to the intracellular pathogen *Legionella pneumophila*. *Curr. Biol.* **13**, 27–36 (2003).
13. Molofsky, A. B. et al. Cytosolic recognition of flagellin by mouse macrophages restricts *Legionella pneumophila* infection. *J. Exp. Med.* **203**, 1093–1104 (2006).
14. Ren, T., Zamboni, D. S., Roy, C. R., Dietrich, W. F. & Vance, R. E. Flagellin-deficient *Legionella* mutants evade caspase-1- and Naip5-mediated macrophage immunity. *PLoS Pathog.* **2**, e18 (2006).
15. Zamboni, D. S. et al. The *Birc1e* cytosolic pattern-recognition receptor contributes to the detection and control of *Legionella pneumophila* infection. *Nature Immunol.* **7**, 318–325 (2006).
16. Fortier, A., de Chastellier, C., Balor, S. & Gros, P. *Birc1e/Naip5* rapidly antagonizes modulation of phagosome maturation by *Legionella pneumophila*. *Cell. Microbiol.* **9**, 910–923 (2007).
17. Lightfield, K. L. et al. Differential requirements for NAIP5 in activation of the NLRC4 inflammasome. *Infect. Immun.* **79**, 1606–1614 (2011).
18. Lamkanfi, M. et al. The Nod-like receptor family member Naip5/Birc1e restricts *Legionella pneumophila* growth independently of caspase-1 activation. *J. Immunol.* **178**, 8022–8027 (2007).
19. Akhter, A. et al. Caspase-7 activation by the Nlr4/Ipaf inflammasome restricts *Legionella pneumophila* infection. *PLoS Pathog.* **5**, e1000361 (2009).
20. Franchi, L. et al. Critical role for Ipaf in *Pseudomonas aeruginosa*-induced caspase-1 activation. *Eur. J. Immunol.* **37**, 3030–3039 (2007).
21. Sutterwala, F. S. et al. Immune recognition of *Pseudomonas aeruginosa* mediated by the IPAF/NLRC4 inflammasome. *J. Exp. Med.* **204**, 3235–3245 (2007).
22. Miao, E. A., Ernst, R. K., Dors, M., Mao, D. P. & Aderem, A. *Pseudomonas aeruginosa* activates caspase 1 through Ipaf. *Proc. Natl Acad. Sci. USA* **105**, 2562–2567 (2008).
23. Suzuki, T. et al. Differential regulation of caspase-1 activation, pyroptosis, and autophagy via Ipaf and ASC in *Shigella*-infected macrophages. *PLoS Pathog.* **3**, e111 (2007).
24. Damiano, J. S., Oliveira, V., Welsh, K. & Reed, J. C. Heterotypic interactions among NACHT domains: implications for regulation of innate immune responses. *Biochem. J.* **381**, 213–219 (2004).
25. Miki, T. et al. Chromobacterium pathogenicity island 1 type III secretion system is a major virulence determinant for *Chromobacterium violaceum*-induced cell death in hepatocytes. *Mol. Microbiol.* **77**, 855–872 (2010).
26. Worrall, L. J., Lameignere, E. & Strynadka, N. C. Structural overview of the bacterial injectisome. *Curr. Opin. Microbiol.* **14**, 3–8 (2011).
27. Poyraz, O. et al. Protein refolding is required for assembly of the type three secretion needle. *Nature Struct. Mol. Biol.* **17**, 788–792 (2010).
28. Hsu, L. C. et al. A NOD2–NALP1 complex mediates caspase-1-dependent IL-1β secretion in response to *Bacillus anthracis* infection and muramyl dipeptide. *Proc. Natl Acad. Sci. USA* **105**, 7803–7808 (2008).
29. Poeck, H. et al. Recognition of RNA virus by RIG-I results in activation of CARD9 and inflammasome signaling for interleukin 1β production. *Nature Immunol.* **11**, 63–69 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank V. Dixit for providing *Nlr4* and *Asc* knockout mice, K. Fitzgerald, D. Radzioch and A. Ding for immortalized macrophages, R. Vance for *Naip5^{ΔVJ}* cDNA, M. Donnenberg and J. Girón for EPEC strains, E. Miao for flagellin-deficient *S. typhimurium* strain, D. Milton and T. Hoang for bacterial vectors and T. Miki for *C. violaceum* strains. We are grateful to C. Yao for helping with flow cytometry, and Y. Xu and the NIBS animal facility for handling mouse lines. We thank members of the F.S. laboratory for helpful discussions and technical assistance. This work was supported by the National Basic Research Program of China (973 Programs, 2010CB835400 and 2012CB518700).

Author Contributions Y.Z. and J.Y. performed experiments, assisted by J.S., Y.-N.G., Q.L., H.X. and L.L. Y.Z., J.Y. and F.S. analysed the data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.S. (shaofeng@nibs.ac.cn).

METHODS

Plasmids, antibodies and reagents. DNAs for flagellin were amplified from the corresponding bacterial genomic DNA, and cloned into pET28a-LFn vector (Addgene) for recombinant expression in *E. coli* as described previously^{30,31}. BsaK and PrgJ DNAs were amplified from *B. thailandensis* E264 and *S. typhimurium* LT2 strains, respectively, and inserted into the same pET28a-LFn vector. DNAs for CprI, CprJ, CorB and CorC were amplified from *C. violaceum* strain (ATCC accession 12472) and also cloned into pET28a-LFn vector to prepare recombinant LFn fusion protein. PA expression plasmid was also obtained from Addgene. To construct the complementation plasmid for the *C. violaceum* mutant, CprI or CprJ DNAs with ribosome binding site (RBS) sequence were cloned into the pBBR1MCS2 vector. Expression plasmids for pro-caspase-1 and pro-IL-1 β were provided by X. Wang (University of Texas Southwestern Medical Center). cDNAs for mouse NAIP1, NAIP2, NAIP5^{C57BL/6}, NAIP6, human NAIP and NLRC4 were amplified from IMAGE EST clones (40130690, 40086453, 6850660, 100068362, 9052275 and 5179909, respectively) and mouse NLRC4 was amplified from reverse-transcribed mouse cDNA. For mammalian expression, cDNAs for all NLR proteins were cloned into modified pCS2 vectors with an N-terminal Myc, HA or Flag epitope tag. All truncations and point mutations were generated by standard molecular biology procedures. All plasmids were verified by DNA sequencing.

Antibodies for caspase-1 and Myc epitopes were obtained from Santa Cruz Biotechnology. Other antibodies used in this study include IL-1 β (3ZD; Biological Resources Branch, National Cancer Institute), HA epitope (Covance) and Flag M2 (Sigma). 293T and HeLa cells obtained from ATCC were grown in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum and 2 mM L-glutamine at 37 °C in a 5% CO₂ incubator. Cell culture products were from Invitrogen and all other chemicals were Sigma-Aldrich products unless noted.

Mouse BMMs and human monocyte-derived macrophages. C57BL/6 wild-type mice were from Vital River Laboratory Animal Technology Co. and caspase-1^{-/-} mice³² were obtained from the Jackson Laboratory. *Nlr4*^{-/-} and *Asc*^{-/-} mice³³ were provided by V. Dixit (Genentech). All knockout alleles have been crossed onto the C57BL/6 background. All animal experiments were conducted following the Ministry of Health national guidelines for housing and care of laboratory animals and performed in accordance with institutional regulations after review and approval by the Institutional Animal Care and Use Committee at National Institute of Biological Sciences. Primary BMMs were prepared by following a standard procedure as previously described³⁴. An immortalized macrophage line derived from C57BL/6 mice was provided by K. A. Fitzgerald (University of Massachusetts Medical School) and TLR4-deficient immortalized BMMs was a gift from A. Ding (Cornell University). Human U937 monocytes obtained from ATCC were cultured in RPMI-1640 containing 10% FBS and 2 mM L-glutamine and grown at 37 °C with 5% CO₂. 50 ng ml⁻¹ PMA was used to induce U937 differentiation for 48 h. Differentiated U937 cells were digested with 2 mM EDTA in PBS and subcultured in 24-well plates for further experiment.

Yeast two-hybrid and co-immunoprecipitation assays. Indicated flagellin, *bsaK* and *prgJ* genes were cloned into the bait vector pLexAde, and mouse *Naip1*, *Naip2*, *Naip5*, *Naip6* cDNAs and *Nlr4* cDNAs were cloned into the prey vector pVP16. The bait and prey plasmids were co-transformed into the reporter *Saccharomyces cerevisiae* strain L40 by using the lithium acetate method. Two-hybrid assays were performed by following a classical procedure³⁵.

For immunoprecipitation, 293T cells were transfected with indicated plasmids. Cells were harvested and lysed in a buffer containing 50 mM Tris-HCl (pH 7.6), 150 mM NaCl and 1% Triton X-100 supplemented with a protease inhibitor mixture (Roche Molecular Biochemicals). Precleared lysates were subjected to anti-Flag M2 immunoprecipitation by following the manufacturer's instructions. The beads were washed three times with the lysis buffer and the immunoprecipitates were eluted in the SDS sample buffer followed by immunoblotting analysis. All the immunoprecipitation assays were performed more than three times and representative results are shown in the figures.

Purification of recombinant proteins. *E. coli* BL21 (DE3) strains harbouring the expression plasmids were grown in Luria-Bertani medium (tryptone, 10 g l⁻¹, yeast extract, 5 g l⁻¹, NaCl, 10.0 g l⁻¹) supplemented with appropriate antibiotics. Protein expression was induced overnight at 22 °C with 0.4 mM isopropyl- β -D-thiogalactopyranoside (IPTG) after OD_{600 nm} reached 0.8. Bacteria were harvested and lysed in a buffer containing 50 mM Tris-HCl (pH 7.6), 300 mM NaCl and 25 mM imidazole. His-tagged proteins were purified by affinity chromatography using Ni-NTA beads (Qiagen). To remove the majority of endotoxin contaminants, proteins bound onto the Ni-NTA column were subjected to an additional wash with 60% isopropanol in the wash buffer (>30 \times column volume). Proteins were then eluted with 250 mM imidazole in 50 mM Tris-HCl (pH 7.6) and 300 mM NaCl. Eluted samples were further dialysed against a buffer containing 50 mM Tris-HCl (pH 7.6) and 150 mM NaCl to remove the imidazole. Protein

concentrations were estimated by Coomassie blue staining of SDS-PAGE gels using BSA as the standards.

NLRC4 inflammasome reconstitution in HeLa and 293T cells. For reconstitution in 293T or HeLa cells, cells were seeded into a 6-well plate 12 h before transfection with indicated combinations of plasmids using the Vigofect reagents (Vigorous). The amounts of plasmids used are 2 μ g for pro-human IL-1 β , 100 ng (HeLa cells) or 50 ng (293T cells) for caspase-1, 100 ng for NLRC4 and 100 ng for NAIP proteins. Twenty-four hours later, LFn-flagellin or LFn-BsaK/PrgJ/CprJ or LFn-CprI together with PA proteins was added into the culture medium at the final concentration of 10 μ g ml⁻¹ for HeLa cells (2 μ g ml⁻¹ for 293T cells) and incubated for another 12 h. Cells were harvested and lysed in a buffer containing 50 mM Tris-HCl (pH 7.6), 150 mM NaCl and 1% Triton X-100. Lysates were resolved onto SDS-PAGE gels followed by anti-IL-1 β immunoblotting analysis. All the reconstitution experiments were performed more than three times and representative results are shown in the figures.

RNAi knockdown. For siRNA knockdown, immortalized BMMs were cultured in 24-well plates at a density of 4 \times 10⁴ per well, and siRNA transfection was performed using the INTERFERin reagent (Polyplus Transfection) by following the manufacturer's instruction. 2 μ l of 20 μ M siRNA (final concentration, 100 nM) and 2 μ l of INTERFERin reagents were used for each well. Sixty hours after transfection, knockdown efficiency and caspase-1 activation were monitored by quantitative real-time PCR (qRT-PCR) and anti-caspase-1 immunoblotting analysis, respectively.

To achieve stable knockdown in immortalized BMMs or U937 cells, shRNAs targeting NAIP5, NAIP2, human NLRC4 or human ASC (listed in Supplementary Table 1) were cloned into a modified lentiviral vector pLKO.1, in which puromycin resistance gene was replaced by GFP coding sequence. pLKO.1-GFP shRNA plasmids were transfected together with two packing plasmids (pCMV-dR8.2 dvpr and pCMV-VSV-G, both from Addgene) into 293T cells. Lentivirus expressing shRNA was collected from the supernatant 48 h after transfection and was used to infect BMMs for another 48 h or U937 cells for 12 h. GFP-positive cells were sorted out by flow cytometry. The pool of sorted cells were either directly used in subsequent functional assays or diluted into 96-well plates to obtain single clones. Knockdown efficiency was examined by qRT-PCR analysis or immunoblotting analysis (for ASC knockdown in U937 cells).

Caspase-1-mediated inflammasome activation assays. To assay caspase-1 activation, culture supernatants of macrophages treated with indicated stimuli were subjected to TCA precipitation and the precipitates were analysed by anti-caspase-1 immunoblotting to detect both pro-Casp1 and processed mature p10 fragment; cell lysates were blotted with Casp1 and actin antibodies to show the level of pro-Casp1 in cell lysates and actin loading, respectively. All caspase-1 activation assays in response to LFn-mediated protein delivery and bacterial infection were repeated at least three times and the representative results are shown in the figures. Mature IL-1 β released into the culture supernatants was measured by using the IL-1 β ELISA kit (Neobioscience Technology Company). Pyroptotic cell death was measured by the lactate dehydrogenase (LDH) assay using CytoTox 96 Non-Radioactive Cytotoxicity Assay kit (Promega). Cell viability was determined by the CellTiter-Glo Luminescent Cell Viability Assay (Promega).

qRT-PCR analysis. For qRT-PCR analysis, total RNA was extracted by TRIzol (Invitrogen) and digested with DNase I (Invitrogen). One microgram of total RNA was reverse-transcribed into cDNA using M-MLV reverse transcriptase (Promega). qRT-PCR analysis was performed using the SYBR Premix Ex Taq (TaKaRa) on Applied Biosystems 7500 Fast Real-Time PCR System. Primers used for qRT-PCR analysis are listed in Supplementary Table 2. The mRNA level of targeted genes was normalized to that of *Gapdh* for mouse BMMs or to that of actin for U937 cells.

Bacterial manipulation and macrophage infection. *L. pneumophila* strains were cultured on buffered charcoal yeast extract agar supplemented with 0.1 mg ml⁻¹ thymidine (BCYET). For infection, bacteria were scraped, diluted in sterile water and added to cells. EPEC strains (E2348/69) were grown overnight in 2 \times YT (tryptone, 16.0 g l⁻¹, yeast extract, 10.0 g l⁻¹, NaCl, 5.0 g l⁻¹) medium without shaking, and then diluted 1:40 in DMEM medium for 4 h to induce the expression of type III secretion system before infection. For *S. typhimurium* infection, overnight 2 \times YT culture was diluted 1:100 and grown for 3 h to induce SPI-1 expression. *B. thailandensis* E264 was obtained from ATCC and cultured as described³¹. Wild-type *C. violaceum* strain (ATCC 12472) was provided by N. Okada and cultured as previously described²⁵. To infect U937 cells, the indicated *C. violaceum* strain cultured overnight at 37 °C in LB broth under conditions of vigorous shaking was diluted 1:100 in fresh LB broth, and further grown for 3 h to obtain an optical density at A600 of 2.0 to 2.5. The bacteria were diluted in serum-free RPMI-1640 medium to achieve a multiplicity of infection (MOI) of 10. All infection experiments were performed with a centrifugation of 1,000g for 10 min at 22 °C.

The flagellin-deficient *L. pneumophila* strain (Lp02 Δ fliA) was generated by standard homologous recombination using the suicide plasmid pSR47 s. Deletion of genes encoding the type III rod protein in EPEC (Δ escI) and *S. typhimurium* (Δ prgJ) strains was achieved by using the suicide vector pCVD442 as described previously³⁶. For gene deletion in *B. thailandensis*, a modified suicide vector pDM4-*pheS* expressing a mutant phenylalanine synthetase (PheS) for counter-selection³⁷ was constructed first. Briefly, the *sacB* gene in the commonly used suicide vector pDM4 (provided by D. Milton and L. Gong) was replaced with a 1.1-kb PS12-*pheS* fragment (PS12, the promoter of the *B. pseudomallei* *rpsL* gene) amplified from pBBR1MCS-Km-*pheS* (provided by T. T. Hoang). A PCR fragment containing flanking sequences of the target gene was then cloned into pDM4-*pheS*. The resulting targeting vector was transferred into *B. thailandensis* through *E. coli* SM10 (λ pir)-mediated conjugational mating. The transconjugants were selected in LB agar medium containing chloramphenicol (50 μ g ml⁻¹) and streptomycin (100 μ g ml⁻¹). The integrants were further screened for markerless in-frame deletion by growth on M9 agar plates supplemented with 20 mM glucose and 0.1% p-chlorophenylalanine. All the mutants were verified by PCR and DNA sequencing. Both flagellin genes in *B. thailandensis* E264, *fliC* (open reading frame (ORF), BTH_I3196) and *fliC2* (ORF, BTH_II0151), were deleted to obtain the flagellin-deficient strain. For gene deletion in *C. violaceum*, the original pDM4-*SacB* suicide vector was used. Briefly, a PCR fragment containing flanking sequence of the targeted gene was cloned into pDM4-*SacB*. The resulting targeting vector was transferred into *C. violaceum* through *E. coli* SM10 (λ pir)-mediated conjugational mating. The transconjugants were selected in LB agar medium containing chloramphenicol (17 μ g ml⁻¹) and nalidixic acid (25 μ g ml⁻¹). The integrants were further screened for markerless in-frame deletion by growth on LB agar plates containing 16% sucrose without NaCl. Detailed information for all deletion strains are listed in Supplementary Table 3. All the mutants were verified by PCR and DNA sequencing.

To examine the role of flagellin in stimulating caspase-1 activation during mouse macrophage infection, wild-type, type III secretion-deficient Δ escN (CVD452,

provided by M. Donnenberg) and flagellin-deficient Δ fliC (AGT01, provided by J. A. Girón) strains of EPEC E2348/69 were used to infect immortalized BMMs at a MOI of 10 for 2 h. Wild-type, type III-deficient Δ bipB and flagellin-deficient Δ fliC/*fliC2* strains of *B. thailandensis* were used to infect J774 mouse macrophages at a MOI of 10 for 2 h. To assay the physiological function of NAIP5 in detecting bacterial flagellin, control or *Naip5* stable knockdown immortalized BMMs were infected with *S. typhimurium* strain (wild type, ATCC 14028 s or Δ fliC Δ fliB mutant, fliC::Tn10 fljB5001::Mud-Cm; both strains were provided by E. A. Miao) for 15 min, or *L. pneumophila* (Lp02 or Lp02 Δ fliA) for 40 min at a MOI of 50. To assay the function of NAIP2 in detecting the type III rod protein during infection, control or *Naip2* stable knockdown immortalized BMMs were infected with *S. typhimurium* (Δ fliC Δ fliB or Δ fliC Δ fliB Δ prgJ) for 30 min or EPEC E2348/69 strain (Δ fliC or Δ fliC Δ escI) for 2 h at a MOI of 50. Supernatants and cell lysates of infected macrophages were collected and subjected to caspase-1 activation assays described above.

30. Yao, Q. *et al.* A bacterial type III effector family uses the papain-like hydrolytic activity to arrest the host cell cycle. *Proc. Natl Acad. Sci. USA* **106**, 3716–3721 (2009).
31. Cui, J. *et al.* Glutamine deamidation and dysfunction of ubiquitin/NEDD8 induced by a bacterial effector family. *Science* **329**, 1215–1218 (2010).
32. Li, P. *et al.* Mice deficient in IL-1 β -converting enzyme are defective in production of mature IL-1 β and resistant to endotoxic shock. *Cell* **80**, 401–411 (1995).
33. Mariathasan, S. *et al.* Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. *Nature* **430**, 213–218 (2004).
34. Boyden, E. D. & Dietrich, W. F. Nalp1b controls mouse macrophage susceptibility to anthrax lethal toxin. *Nature Genet.* **38**, 240–244 (2006).
35. Vojtek, A. B. & Cooper, J. A. Rho family members: activators of MAP kinase cascades. *Cell* **82**, 527–529 (1995).
36. Dong, N., Liu, L. & Shao, F. A bacterial effector targets host DH-PH domain RhoGEFs and antagonizes macrophage phagocytosis. *EMBO J.* **29**, 1363–1376 (2010).
37. Barrett, A. R. *et al.* Genetic tools for allelic replacement in *Burkholderia* species. *Appl. Environ. Microbiol.* **74**, 4498–4508 (2008).